

# UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS  
Departamento de Bioquímica y Biología Molecular I



## TESIS DOCTORAL

**Biología estructural de dos nuevas familias de proteínas de superficie  
bacteriana implicadas en infección y en regulación de la pared**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Noelia Bernardo García**

Director

Juan Antonio Hermoso Domínguez

**Madrid, 2015**



---

**Biología estructural de dos nuevas familias de  
proteínas de superficie bacteriana implicadas en  
infección y en regulación de la pared.**

---

Noelia Bernardo García

*Universidad Complutense de Madrid*  
*Consejo Superior de Investigaciones Científicas*



**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I**



**TESIS DOCTORAL**

**Biología estructural de dos nuevas familias de proteínas de superficie bacteriana implicadas en infección y en regulación de la pared.**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Noelia Bernardo García**

Director

Juan Antonio Hermoso Domínguez

**Madrid, 2015**



---

## RESUMEN

Las bacterias han desarrollado múltiples estrategias de resistencia frente a antibióticos, convirtiéndose en una grave amenaza para la salud pública. Muchas de ellas presentan resistencia a más de un antibiótico y son capaces de sobrevivir a los tratamientos actuales. Por otro lado, el coste económico para la salud pública derivado de la resistencia a antibióticos es altísimo. Como consecuencia de todo ello, la efectividad de los antibióticos es cada vez menor y se hace necesario buscar nuevas alternativas terapéuticas.

Hasta el momento la diana por excelencia para combatir las infecciones bacterianas es la ruta de biosíntesis del peptidoglicano. Además, en la superficie de la bacteria encontramos proteínas que son consideradas potenciales factores de virulencia por su contribución en las interacciones hospedador-patógeno. En este trabajo se estudian dos familias de proteínas presentes en la superficie bacteriana implicadas en la regulación de la biosíntesis de la pared y en la interacción con proteínas de la matriz extracelular del hospedador.

En *Vibrio cholerae*, los D-aminoácidos no canónicos (NCDAAs) tienen un papel regulador en la síntesis del peptidoglicano: influyen en su composición, cantidad y resistencia a través de su incorporación directa en el peptidoglicano o a través de la regulación de las enzimas que intervienen en su síntesis y modificación. Además los D-aminoácidos intervienen en procesos celulares clave para la bacteria como son el crecimiento bacteriano, la germinación y la formación del biofilm.

Las racemasas son las enzimas implicadas en la conversión de los L-aminoácidos en D-aminoácidos. En este trabajo se presentan las características estructurales que definen tanto a las alanina racemasas como a las racemasas de amplio espectro. El primer representante de las racemasas de amplio espectro es la racemasa de *V. cholerae* BsrV, enzima periplásmica capaz de racemizar diez de los diecinueve aminoácidos naturales. La resolución de la estructura tridimensional de BsrV ha permitido establecer los determinantes estructurales que hacen posible la plasticidad del mecanismo catalítico de la enzima.

La superficie bacteriana está compuesta por numerosas proteínas que son consideradas factores de virulencia. En *Streptococcus pneumoniae* existe un grupo de proteínas que no tiene un motivo de anclaje a membrana o un péptido líder, a estas proteínas se las ha clasificado como proteínas no clásicas de superficie (NCSP). Son proteínas que, en el citoplasma, llevan a cabo una función relacionada con el metabolismo celular y que, cuando se encuentran localizadas en la superficie celular, pueden realizar diversas funciones, actuando como adhesinas. *S. pneumoniae* es capaz de unir plasminógeno humano a través de las proteínas de superficie, el reclutamiento del mismo en la superficie bacteriana promueve su conversión en la proteasa plasmina, lo que facilita la disolución de las cadenas de fibrina y la degradación de la matriz extracelular del hospedador.

---

La proteína fosfoglicerato quinasa (PGK) de neumococo ha sido identificada como una proteína no clásica de superficie. En este trabajo se ha demostrado que dicha enzima está presente en la superficie celular y que se encuentra localizada junto con el plasminógeno. La estructura tridimensional de la PGK ha permitido identificar el sitio de unión a plasminógeno, localizado en la cara opuesta al sitio catalítico. Además, una propiedad única de la PGK es la capacidad de reclutar el activador de plasminógeno tPA promoviendo la activación del plasminógeno y la consecuente degradación de las cadenas de fibrina, facilitando la dispersión del patógeno dentro del hospedador.

La caracterización estructural de ambas familias de proteínas de superficie sentará las bases para la búsqueda de nuevas dianas para antibióticos, de manera que, cuando estas proteínas se inhiban, la capacidad del patógeno para desarrollar resistencia durante la infección se vería afectada gravemente.

---

## SUMMARY

Bacteria can be found in many different habitats; some of them can grow under severe conditions including acidic pHs, high temperature, radioactive residues, etc. Inside the human body there is a high amount of bacteria concentrated on the skin and in the gastrointestinal tract, playing a positive role for the human. The presence of different physical and chemical barriers in combination with the immune system prevents the entrance of infectious microorganisms into the host organism. However, some infectious agents could evade the immune response.

Bacterial pathogenesis and virulence are related with the ability of the pathogen to colonize and invade host tissues, which results in the development of resistance to the host immune response. Some factors that allow pathogens to evade the immune response are related with the bacterial cell wall and capsule.

*Streptococcus pneumoniae* is a capsulated Gram-positive bacterium that colonizes the upper respiratory tract of humans and mediates local infections, e.g. otitis media. More importantly, *S. pneumoniae* is known to cause severe systemic diseases such as pneumonia and meningitis in children, elderly patients, and immune compromised people. The development of antibiotic resistance, results in the high mortality and morbidity rates of pneumococcal infections and is the primary reason for the current intensive efforts in vaccine-based prevention strategies. *Vibrio cholerae* is a Gram negative pathogen that produces cholera toxin when entering the intestines to cause infection. Another mechanism used by bacteria to evade the host immune response is the presence of different cellular structures, which is the case of the pili in *Escherichia coli* that enable the bacterium to adhere to host epithelial cells.

Here, we also referred to non-pathogenic organisms like *Bacillus subtilis*, which is present in the soil and is considered a model organism for the study of endospore formation. Bacterial sporulation is an adaptive response to environmental stress, such as starvation, and involves a finely controlled developmental program. During sporulation this organism can produce antibiotics. In addition, *B. subtilis* is able to form biofilms, this cell community adaptation could be positive for the neighbouring organisms, e.g. in plants biofilm formation around the radicular system prevents infection by other pathogenic microorganism and promotes plant growth.

Peptidoglycan is the main structural component of the bacterial cell wall and nowadays, the most successful target for antibacterial agents is the peptidoglycan biosynthetic pathway. Additionally, all of the enzymes involved in the peptidoglycan biosynthesis are essential and conserved in both Gram-positive and Gram-negative bacteria, and most of them do not have homologues in mammals. For these reasons, these enzymes are crucial for the development of new antibiotics.

---

Peptidoglycan imparts to the cell its shape, strength, and resistance to osmotic pressure. PG is found on the outside of the cytoplasmic membrane of almost all bacteria and serves as a scaffold for anchoring cell envelope components. It is composed of linear glycan strands containing a repeating disaccharide unit of *N*-acetyl glucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc). These glycan strands are cross-linked by short peptides. In Gram-negative bacteria, a single layer of PG, which is found in the periplasmic space between the inner and outer cell membranes, is sufficient to maintain the cell's mechanical stability. In Gram-positive bacteria, which lack an outer cell membrane, the cell wall is thicker, consisting of many layers of PG.

The biosynthesis of the PG is divided into three main stages. PG biosynthesis starts in the cytoplasm, where the UDP-MurNAc-pentapeptide and UDP-GlcNAc are synthesized. Second, bactoprenol-linked UDP-disaccharide pentapeptide (lipid II) is formed to flip PG precursors across the membrane. Third, PG precursors are polymerized by the transglycosylase and transpeptidase activities of the synthetic penicillin-binding proteins (PBPs) resulting in the murein sacculus. Transpeptidase enzymes are natural targets for  $\beta$ -lactam antibiotics, e.g. penicillins and cephalosporins.

An extensive structural analysis of two new families of proteins located on the bacterial surface has been carried out with the aim to elucidate their implications in cell-wall regulation and virulence.

D-amino acids (DAA) are key constituents of the short peptides that cross-link the glycan chains within the PG polymer. D-amino acids are mirror images of L-amino acids, in which the chirality at the alpha carbon has been inverted. They are highly resistant to protease-mediated degradation and have low immunogenicity. These findings and the fact that D-amino acids have recently emerged as extracellular effectors with remarkable roles in microbiology (e.g. cell wall regulators, sporulation, biofilm, interspecies communication...) make them attractive candidates for peptidomimetics drug design.

Recently, it has been demonstrated that diverse bacteria produce and release to the extracellular medium high amounts of DAA that are not canonical components of the cell wall (e.g. D-Ala and D-Glu). In *Vibrio cholerae*, non-canonical DAA (NCDAA) edit PG architecture to control the amount and strength of PG during the stationary phase of cell growth. In addition to the cell wall, DAA are directly involved in important biological processes taking place in bacteria, such as bacterial growth, spore germination, and biofilm formation. DAA can break up the biofilm previously arranged. Furthermore, D-Ala has been reported to inhibit spore germination.

Racemases are the specific enzymes that catalyze the interconversion between L-amino acids and D-amino acids. These enzymes can be classified into two main classes depending on whether their mode of action is dependent or independent of the coenzyme pyridoxal-5-phosphate (the biologically active form of vitamin B<sub>6</sub>). Pyridoxal-5-phosphate (PLP) dependent racemases in eukaryotes include aspartate



---

racemase (AspR), alanine racemase (AlaR) and serine racemase (SerR). The prokaryotic PLP dependent racemases include AlaR, SerR, ArgR and recently, a new class of broad-spectrum racemases (BsrR). Among the PLP-independent prokaryotic enzymes are aspartate racemases (AspR), glutamate racemases (GluR) and proline racemases (ProR). Both PLP-dependent and PLP-independent types of enzymes deprotonate the  $\alpha$ -carbon of the L-amino acid to reprotonate it on the opposite side, thus generating the D-form of the amino acid.

Here, we discovered the structure and catalytic mechanism and the structural features of the new family of broad-spectrum racemases. The first family member of this new family is the broad-spectrum racemase from *Vibrio cholerae* (BsrV). BsrV is a periplasmic enzyme that racemizes ten amino acids (Ala, Leu, Met, Ser, Cys, Gln, Asn, His, Lys and Arg) of the nineteen natural chiral amino acids. Interestingly, BsrV can interconvert radically different residues such as aliphatic (non- $\beta$ -branched) and positively charged residues. However, negatively charged, aromatic or  $\beta$ -branched aliphatic (Ile, Val and Thr) amino acids are not substrates of BsrV. Comparative structural analyses of the BsrV-family of broad-spectrum racemases and other more restricted enzymes (AlaR) suggest that the catalytic plasticity of these racemases has three essential determinants: the arrangement and features of the entry site, the width of the channel leading to the active site, and its catalytic environment, which is able to accommodate and stabilize substrates of diverse chemical natures. The wider entry site and channel of BsrV are likely to facilitate interactions with amino acid substrates larger than alanine. Indeed, structural, biochemical, and modeling analyses suggest that the entry channel of BsrV can also accommodate peptide chains that are markedly larger. A molecular footprint of conserved residues that mediate the multispecificity of BsrV has been identified through bioinformatics tools and validated experimentally, this molecular signature will be helpful in identifying a large family of BsrV-like multi-specific racemases in bacteria.

Considering the relevant role of NCDAA in cell-wall regulation and bacterial metabolism, this new information is paving the way to the design of novel and more effective antibacterials.

Bacterial surface proteins are considered as potential virulence factors because of their contribution to the host-pathogen interactions. Three main groups of surface proteins have been identified in *S. pneumoniae*: (i) around 50 lipoproteins, (ii) up to 18 LPXTG consensus sequence-carrying proteins that are covalently linked via sortase to the cell wall PG, and (iii) up to 16 choline-binding proteins (CBPs). In addition, the cell wall envelope of pneumococci is decorated with another cluster of proteins that lack the classic leader peptide and membrane-anchoring motifs. These proteins are termed as non-classical surface proteins (NCSP) and recent results indicate that they could also play an important role in subverting the physiological function of host-derived proteins.

---

These NCSP proteins are described primarily as cytoplasmic proteins with intracellular roles that are involved in host-pathogen interactions. However, once they are located by unknown mechanisms on the cell surface, they display moon-lighting functions and frequently act as adhesins. Binding of NCSP to host molecules, like fibronectin (PavA) or plasminogen (enolase), promotes pneumococcal invasion and the spread of the infection.

Some bacterial metabolic enzymes, such as enolase and glyceraldehyde-3-phospho dehydrogenase (GAPDH) have been characterized as NCSP proteins. These proteins exhibit moon-lighting functions including the adherence to various host components (e.g. fibronectin and laminin) and the capability to serve as receptors for plasminogen (PLG) recruitment. PLG is a pro-enzyme of the broad-spectrum protease plasmin, which counteracts the blood coagulation cascade. It is composed of a pre-activation peptide, five kringle structures (K1-5) and a serine protease domain. The first four kringle domains of PLG act as inhibitory molecules for angiogenesis and are therefore termed angiostatin, whereas the last kringle domain together with the enzymatic domain are known as mini-plasminogen (mini-PLG). PLG is converted into plasmin by eukaryotic plasminogen activators (urokinase-type plasminogen, uPA and tissue-type plasminogen activator, tPA). These activators initiate the conversion of PLG to plasmin, which hydrolyses polymerized fibrin strands into soluble fibrin degradation products.

The phosphoglycerate kinase (PGK) was here identified as an additional glycolytic enzyme bearing moon-lighting function on the surface of *S. agalactiae*. The pneumococcal PGK is a monomeric two-domain enzyme that mediates transfer of an anhydride phosphate of 1,3-biphospho-D-glycerate (1,3-BPG) to Adenosin 5'-diphosphate magnesium salt (MgADP), yielding 3-phospho-D-glycerate (3-PGA) and Adenosin 5'-triphosphate magnesium salt (MgATP) within glycolytic metabolism. The present thesis provides insights into the moon-lighting functions of pneumococcal PGK by biophysical and structural analyses.

In this work, we demonstrate that PGK is indeed exposed on the pneumococcal surface, colocalising with plasminogen. We provide kinetic evaluation of interactions with human and murine plasminogen, detect two linear plasminogen-binding sites, and localize the crucial amino acids of PGK contributing to plasminogen binding. The high-resolution crystal structure of pneumococcal PGK revealed localization of the plasminogen-binding sites in one surface exposed region of the N-terminal domain and determined that angiostatin is the only domain required for the interaction with PGK. A remarkable and unique property of PGK is the recruitment of the tissue plasminogen activator (tPA), which enhances the activation of plasminogen and the subsequent degradation of fibrin thrombi. This kringle-based interaction with PGK and tPA was predicted by structural modeling and confirmed experimentally by binding studies and plasmin activity assays.

---

*Streptococcus pneumoniae* binds human plasminogen via several surface-exposed binding proteins, but does not express an endogenous plasminogen activator. The recruitment of plasminogen to the pneumococcal surface promotes conversion of plasminogen into proteolytic active plasmin. This leads to the dissolution of fibrin thrombi and to the degradation of host extracellular-matrix proteins.

Considering that fibrinolysis also contributes to complex inflammatory processes, our results describe a new scenario in which this important human pathogen can subvert the host fibrinolytic machinery. During the initial phase of bacterial colonization, this process enhances both dissemination through tissue barriers and entry into the vasculature.

In this work we summarized the important roles of two new protein families located on the bacterial surface: the role of NCDAA in cell-wall regulation bacterial metabolism, and the role of NCSP to recruit host extracellular matrix proteins. Structural and regulatory characterization of both families of proteins will reveal novel drug targets that when inhibited would challenge the adaptability of pathogenic bacteria to develop resistance in adverse conditions (*i.e.* during infection).



# ÍNDICE

<b>RESUMEN</b>	<b>I</b>
<b>SUMMARY</b>	<b>III</b>
<b>ABREVIATURAS</b>	<b>3</b>
<b>1. INTRODUCCIÓN</b>	<b>5</b>
1.1. LA PARED CELULAR BACTERIANA	9
1.2. BIOSÍNTESIS DE PEPTIDOGLICANO	12
1.3. PROTEÍNAS DE SUPERFICIE	14
1.3.1 UNA NUEVA FAMILIA: LAS PROTEÍNAS NCSP	14
1.4. RELEVANCIA DE LOS d-AMINOÁCIDOS EN LA FISIOLÓGIA BACTERIANA	17
1.4.1. DESENSAMBLAJE DEL BIOFILM	20
1.4.2. EFECTO DE LOS d-AMINOÁCIDOS EN ESPORULACIÓN	21
1.5. LAS RACEMASAS, ENZIMAS IMPLICADAS EN LA PRODUCCIÓN DE DAA	23
1.5.1. ALANINA RACEMASAS	24
<b>2. OBJETIVOS</b>	<b>27</b>
<b>3. TÉCNICAS CRISTALOGRÁFICAS</b>	<b>31</b>
3.1. VISIÓN GENERAL DEL PROCESO	33
3.2. CRISTALIZACIÓN	33
3.2.1. CRISTAL Y SIMETRÍA	34
3.2.2. FACTORES QUE AFECTAN A LA CRISTALIZACIÓN	35
3.2.3. BÚSQUEDA DE CONDICIONES DE CRISTALIZACIÓN	37
3.2.4. EXPERIMENTOS DE CRISTALIZACIÓN	38
3.2.5. CRISTALIZACIÓN DE COMPLEJOS DE PROTEÍNA-PROTEÍNA	39
3.2.6. PREPARACIÓN DE LAS MUESTRAS	39
3.3. DIFRACCIÓN DE RAYOS X Y RESOLUCIÓN ESTRUCTURAL	40
3.3.1. FUENTES DE RADIACIÓN	41
3.3.2. PROCESADO DE DATOS DE DIFRACCIÓN	43
3.3.3. MÉTODOS DE RESOLUCIÓN ESTRUCTURAL	45
3.3.4. REFINAMIENTO DE LOS MODELOS.	47
3.3.5. MODELADO MANUAL Y VALIDACIÓN DEL MODELO FINAL.	48
3.4. OTROS PROGRAMAS UTILIZADOS	49
<b>4. TRABAJOS DE INVESTIGACIÓN</b>	<b>51</b>
4.1. INTRODUCCIÓN A LA CRISTALIZACIÓN DE LA FOSFOGLICERATO QUINASA Y APORTACIÓN DEL AUTOR	53
Crystallization and preliminary X-ray diffraction analysis of phosphoglycerate kinase from <i>Streptococcus pneumoniae</i>	57
4.1.1. DISCUSIÓN	65
4.2. INTRODUCCIÓN A LA FOSFOGLICERATO QUINASA Y APORTACIÓN DEL AUTOR	67
Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator.	71
4.2.1. DISCUSIÓN	95
4.3. INTRODUCCIÓN A LA RACEMASA DE AMPLIO ESPECTRO BsrV Y APORTACIÓN DEL AUTOR	97
Structural basis for the broad specificity of a new family of amino-acid racemases.	101
4.3.1. DISCUSIÓN	147
4.4. INTRODUCCIÓN A LA BIOINFORMÁTICA DE RACEMASAS DE AMPLIO ESPECTRO Y APORTACIÓN DEL AUTOR	149

Structural bioinformatics in broad-spectrum racemases: a new path in antimicrobial research.	153
4.4.1. DISCUSIÓN	183
<b>5. MATERIAL INÉDITO</b>	<b>185</b>
<b>5.1. ANEXO I. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA MONOESPECÍFICA DE <i>Bacillus subtilis</i> Bsu17640</b>	<b>187</b>
I.I. CRISTALIZACIÓN	187
I.II. DETERMINACIÓN ESTRUCTURAL	189
I.III. ANÁLISIS ESTRUCTURAL	191
<b>5.2. ANEXO II. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA DE AMPLIO ESPECTRO DE <i>Acinetobacter baumannii</i> BsrAb</b>	<b>193</b>
II.I. CRISTALIZACIÓN	193
II.II. DETERMINACIÓN ESTRUCTURAL	194
II.III. ANÁLISIS ESTRUCTURAL	197
II.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS	197
<b>5.3. ANEXO III. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA DE AMPLIO ESPECTRO DE <i>Kingella oralis</i> BsrKo</b>	<b>201</b>
III.I. CRISTALIZACIÓN	201
III.II. DETERMINACIÓN ESTRUCTURAL	202
III. III. ANÁLISIS ESTRUCTURAL	204
III.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS	205
<b>5.4. ANEXO IV. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA L,D-TRANSPEPTIDASA DE <i>E.coli</i> YnhG</b>	<b>209</b>
IV.I. CRISTALIZACIÓN	209
IV.II. DETERMINACIÓN ESTRUCTURAL	210
IV. III. ANÁLISIS ESTRUCTURAL	212
IV.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS	213
<b>6. CONSIDERACIONES FINALES</b>	<b>215</b>
<b>6.1. EL PAPEL MULTIFUNCIONAL DE LA FOSFOGLICERATO QUINASA DE NEUMOCOCO Y SU IMPLICACIÓN EN VIRULENCIA</b>	<b>217</b>
<b>6.2. MAQUINARIA CATALÍTICA E IMPLICACIONES <i>IN VIVO</i> DE UNA NUEVA FAMILIA DE RACEMASAS DE AMPLIO ESPECTRO.</b>	<b>219</b>
<b>7. CONCLUSIONES</b>	<b>221</b>
<b>8. BIBLIOGRAFÍA</b>	<b>225</b>

---

## ABREVIATURAS

ADP: Adenosín difosfato

AlaR: alanina racemasas

AMP-PNP: adenilil-imidodifosfato

APBS: *Adaptive Poisson-Boltzmann Software*

ARN: ácido ribonucleico

ATP: Adenosín trifosfato

ATTGal: 2-acetoamido-4-amino-2,4,6-tridesoxi-D-galactosa

BsR: Racemasas de amplio espectro (*Broad spectrum racemases*)

BPG: 1,3-bifosfoglicerato

CBPs: Proteínas de unión a colina (*Choline Binding Proteins*)

CDC: Centro de prevención y control de enfermedades (*Centers for Disease Control and Prevention*)

ChoP: fosforilcolina

DAA: D-aminoácidos

DAP: diaminopimélico

ESRF: Sincrotrón europeo (*European Synchrotron Radiation Facility*)

GAPDH: *glyceraldehyde-3-phospho dehydrogenase*

HMM: Alto peso molecular (*High Molecular Mass*)

LLP: *N'-pyridoxyl-lysine-5'-monophosphate*

LMM: Bajo peso molecular (*Low Molecular Mass*)

KCX: *N-carboxylated lysine*

LPS: Lipopolisacárido

LTA: Ácidos lipoteicoicos (*lipoteichoic acids*)

MIB tampón: Malonato, imidazol y ácido bórico

MPD: 2-Metil-2, 4-pentanediol

NAG: N-acetilglucosamina

NAM: N-acetilmurámico

NCDAAs: D Aminoácidos no canónicos (*non-canonical D-aminoacids*)

NCSP: Proteínas no clásicas de superficie (*Non-classical surface proteins*)

NRP: Péptido no ribosomal (*Non Ribosomal Peptide*)

---

PBP: Proteínas de unión a penicilina (*Penicillin Binding Protein*)

PCT: Test de pre-cristalización (*Pre-Crystallization Test*)

PDB: Banco de datos de proteínas (*Protein Data Bank*)

PEG: Polietilenglicol

PG: peptidoglicano

PGK: Fosfoglicerato quinasa (*Phosphoglycerate Kinase*)

PLG: plasminógeno

PLP: Piridoxal-5-fosfato

TA: Ácidos teicoicos (*teichoic acids*)

tPA: *tissue-type plasminogen activator*

TRIS: tris(hidroximetil)aminometano

UDP: uridina difosfato

UP: undecaprenil-fosfato

uPA: *urokinase-type plasminogen activator*

SCN: tiocianato o sulfocianuro

SPG tampón: ácido succínico, fosfato sódico dihidrógeno y glicina (*Succinic acid, Sodium dihydrogen phosphate and glycine buffer*)

SLS: *Swiss Light Source*

3PG: 3-fosfoglicerato (*3-phosphoglycerate*)

$\gamma$ DPGA: ácido poli- $\gamma$ -D-glutámico



## 1. INTRODUCCIÓN



Las bacterias son microorganismos procariotas de unos pocos micrómetros que presentan una gran variedad de morfologías. A diferencia de las células eucariotas, las células procariotas presentan una estructura interna más simple sin núcleo definido y carecen de orgánulos rodeados por membranas.

Las bacterias son los organismos más abundantes en el planeta, se hallan presentes en una gran variedad de ambientes y se encuentran tanto en hábitats acuáticos, como aéreos y terrestres; crecen en las condiciones más extremas tales como manantiales de aguas calientes y ácidas, en desechos radioactivos, en las profundidades del mar y de la corteza terrestre. Algunas, incluso, pueden sobrevivir a las condiciones extremas del espacio exterior.

Las bacterias son imprescindibles para el reciclaje de los elementos, muchos pasos de los ciclos biogeoquímicos dependen de ellas, como por ejemplo la fijación del nitrógeno atmosférico. En el cuerpo humano hay aproximadamente diez veces más bacterias que células humanas, con una gran cantidad de bacterias en la piel y en el tracto digestivo. La mayoría de los microorganismos infecciosos no logran penetrar en el individuo gracias a las barreras físicas y químicas que éste presenta. La barrera física más importante es la piel; la integridad de ésta, junto a la segregación de mediadores químicos, evita el ingreso de los microorganismos patógenos. Asimismo, las mucosas presentan una serie de atributos (secreciones, flujo ciliar) que dificultan el ingreso de microorganismos por esta vía. Además, la existencia de poblaciones microbianas no patógenas residentes (flora normal), también impide la colonización de mucosas por agentes infecciosos. La mayoría de los microorganismos que logran evadir estas barreras y producir infección, son destruidos en pocas horas por mecanismos no específicos de inducción rápida (inmunidad innata). Sin embargo, si un agente infeccioso es capaz de superar esas primeras líneas de defensa, se activará, en la mayoría de los casos, un tipo de respuesta de defensa (inmunidad adaptativa), altamente especializada y específica. Ésta logrará, en la mayoría de las situaciones, controlar la infección y suprimir la enfermedad. Además, de este proceso resultará la generación de memoria inmunológica, que permitirá al individuo en el próximo contacto con el mismo agente, responder con mayor rapidez y eficacia.

La virulencia o patogenicidad de las bacterias se relaciona con atributos que favorecen la colonización e invasión de los tejidos del huésped y que les permiten resistir a la acción del sistema inmune. Aquellos incluyen proteínas de superficie bacteriana con propiedades de adhesinas, mecanismos antifagocitarios e inhibición del complemento o inactivación de sus productos.

Las cápsulas de muchas bacterias confieren resistencia a la fagocitosis y, además, algunas tienen residuos de ácido siálico que inhiben la activación de la vía del complemento (Severi *et al.*, 2007). Por lo tanto, la producción de la cápsula constituye un mecanismo importante de evasión inmune y las bacterias encapsuladas son más virulentas que las que carecen de tal estructura. Muchas bacterias que causan

meningitis (*meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*) son capsuladas. Las cápsulas son un elemento común y esencial de estas bacterias, que les permite sobrevivir en la sangre.

Una de las bacterias objeto de estudio de esta tesis es el microorganismo encapsulado *S. pneumoniae*. El neumococo es una bacteria Gram-positiva que coloniza las vías respiratorias y también produce infecciones locales, como la otitis media. Se sabe que *S. pneumoniae* es el causante de neumonía, septicemia y meningitis en niños, ancianos y pacientes inmunodeprimidos (Cartwright, 2002). A pesar de la gran cantidad de antibióticos disponible, la mortalidad y morbilidad causadas por este patógeno continúan siendo altas.

La producción de toxinas es otro de los mecanismos empleados por bacterias Gram-negativas (por ejemplo *Vibrio cholerae*) para causar enfermedad. Las toxinas son potentes estimuladores de la producción de citoquinas y activador de los macrófagos. Muchas endotoxinas son primariamente citotóxicas, pudiendo destruir por distintos mecanismos a las células a las que se fijan. Durante el proceso infeccioso *V. cholerae* secreta la toxina colérica, una proteína que causa diarrea profusa. Además, la colonización del intestino delgado también requiere de un apéndice filamentoso, flexible y delgado que se encuentra en la superficie bacteriana. Uno de los serotipos de *V. cholerae* presenta también una cápsula polisacáridica. Una característica singular que presenta *V. cholerae* es la percepción de quórum (en inglés *quorum sensing*), un mecanismo de regulación de la expresión genética en respuesta a la densidad de la población celular (Miller & Bassler, 2001).

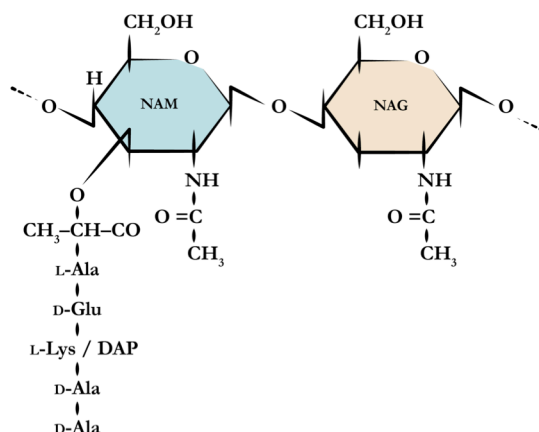
Otro mecanismo usado por las bacterias para evadir la respuesta inmune adaptativa, es la variación genética de antígenos de superficie. Muchas bacterias como *Escherichia coli* y *Neisseria gonorrhoeae*, presentan pili (estructuras implicadas en la adhesión bacteriana a las células del huésped, constituidos por la proteína pilina). Este mecanismo de variación antigénica ayuda a la bacteria a “escapar” de los anticuerpos específicos, aunque el significado principal para el microorganismo sea seleccionar el pili que mejor se adhiera a las células epiteliales del huésped.

Dentro de las bacterias que se presentan en esta tesis se encuentran también organismos no patógenos como *Bacillus subtilis* que está presente en el suelo y es modelo de estudio de la formación de endosporas en bacteria. La formación de esporas tiene lugar en condiciones de escasez de nutrientes, estos organismos pueden producir antibióticos durante la esporulación como la polimixina y la subtilina. Además, *B. subtilis* es capaz de formar biofilms. Éstos pueden resultar beneficiosos para las plantas con las que establecen una interacción mutualista en la que se previene la infección por otro patógeno; la formación de biofilms en el rizoma promueve el crecimiento y sirve como un biosensor. Se emplea en agricultura como fungicida debido a que coloniza el sistema radicular de la planta compitiendo con hongos causantes de enfermedades, lo que le otorga un papel importante en la industria biotecnológica.

Una de las dianas de acción de los agentes antibacterianos con más éxito a nivel clínico y económico, hasta ahora, es la ruta de biosíntesis del peptidoglicano, el principal componente estructural de la pared bacteriana común a la mayoría de las bacterias patógenas. Las enzimas responsables de la biosíntesis de la pared bacteriana son esenciales y muchas de ellas no tienen homólogos en mamíferos. Este hecho, y que estén presentes tanto en patógenos Gram-positivos como Gram-negativos, las convierten en objeto de estudio para el desarrollo de antibióticos de amplio espectro. Las bacterias han desarrollado resistencias frente a los antibióticos empleados para combatirlas y la búsqueda de nuevas dianas de actuación para el desarrollo de fármacos se hace indispensable.

### 1.1. LA PARED CELULAR BACTERIANA

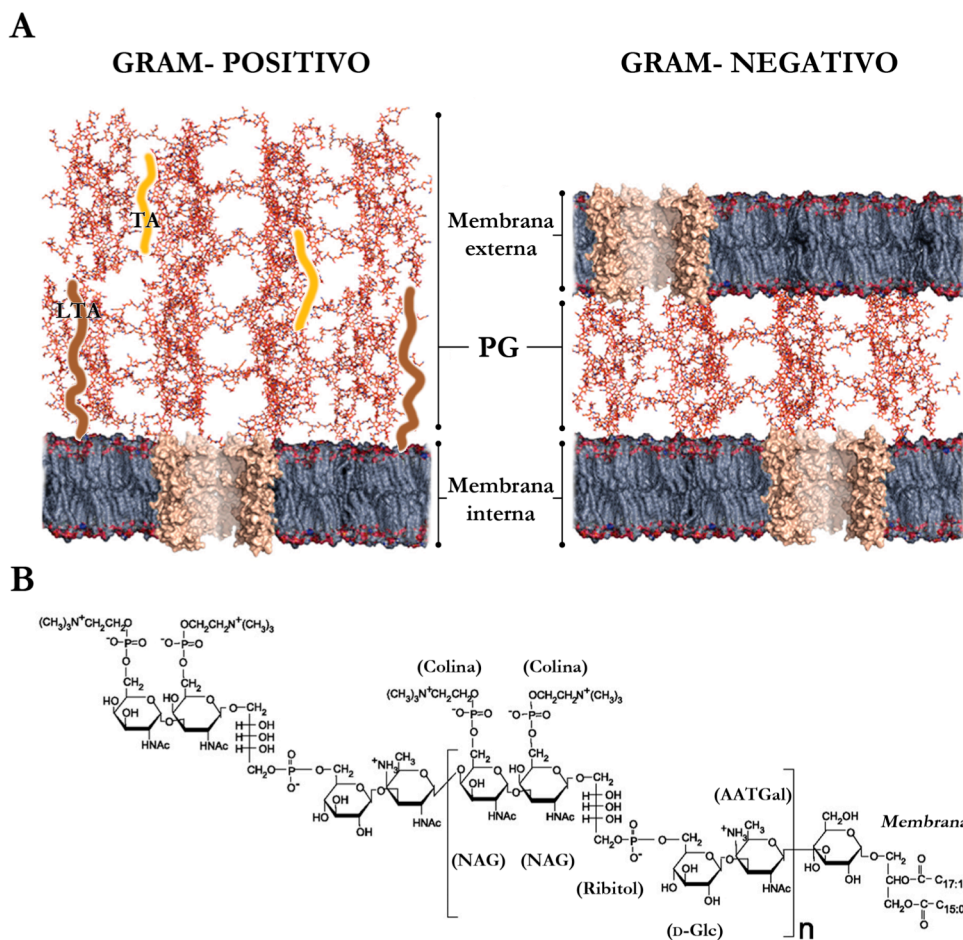
La pared celular bacteriana da integridad estructural a la célula, su principal función es la de proteger a la célula de la presión de turgor, presión intracelular producida por la concentración elevada de proteínas y otras moléculas comparada con el exterior.



**Figura 1.1. Esquema de las unidades esenciales que componen el peptidoglicano de bacterias:** NAG, *N*-Acetilglucosamina y NAM, *N*-Acilmurámico.

El principal componente de la pared celular es el peptidoglicano (PG) que, situado alrededor de la membrana citoplasmática, es responsable de dar rigidez a la pared celular y determinar la forma de la bacteria. La capa de PG está compuesta por una red polimérica elaborada en la que se van alternando unidades de *N*-acetilglucosamina (NAG) y ácido *N*-acilmurámico (Kaku *et al.*). Estos azúcares están conectados por enlaces  $\beta$ -(1-4)-glicosídicos. Cada unidad de NAM tiene conectada una cadena peptídica, la cual puede contener de tres a cinco aminoácidos. Dependiendo del tipo de bacteria la secuencia de aminoácidos que está unida al ácido *N*-acilmurámico puede variar (Fig. 1.1).

Las bacterias se pueden clasificar en función de la composición de la pared celular en Gram-positivas y Gram-negativas. La distinción inicial entre estos dos tipos se llevó a cabo mediante una tinción diferencial llamada tinción de Gram; pero la diferente reacción a la tinción de Gram se basa en las diferencias existentes en la estructura de las paredes celulares. La morfología de las paredes celulares es muy distinta en las células Gram-positivas y Gram-negativas (Fig. 1.2 a). La pared de las Gram-negativas está compuesta por dos membranas y una capa de peptidoglicano, mientras que la pared de las Gram-positivas está formada fundamentalmente por una membrana y una espesa capa de PG.



**Figura 1.2. Pared celular de organismos Gram-positivos y Gram-negativos. a)** Diferencias estructurales relativas a la pared celular entre bacterias Gram-positivas y Gram-negativas. **b)** Modelo de ácido lipoteicoico en el que se observan las repeticiones de d-GalNac(ChoP)-D-GalNac(ChoP)-ribitol-5-P-6-D-Glc-AATGal y la unión a la membrana mediante una unidad de D-Glc.

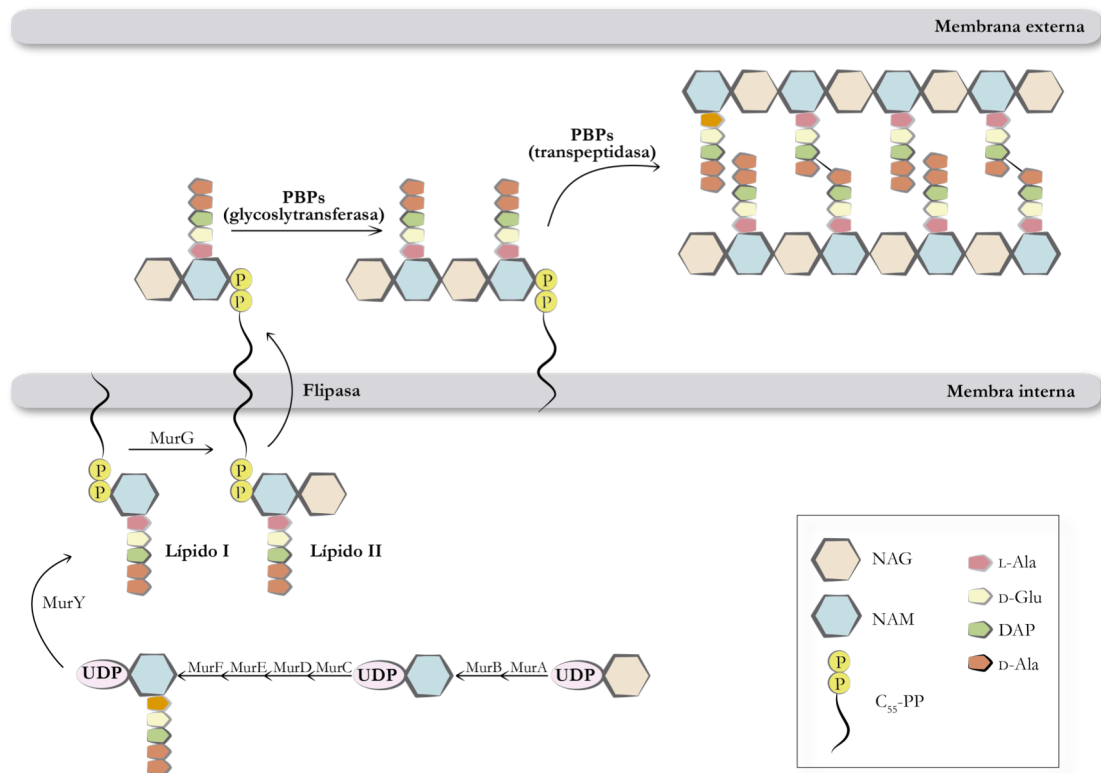
En las bacterias Gram-positivas (*S. pneumoniae*) el PG representa el 90% de la pared. La pared celular de neumococo además presenta ácidos teicoicos (TA, *teichoic acids*) unidos covalentemente al peptidoglicano y ácidos lipoteicoicos (LTA, *lipoteichoic acids*) anclados a la membrana. En neumococo las unidades de TA y LTA tienen una estructura química idéntica (Vollmer, 2007) (Fig. 1.2 b). Las unidades

repetidas de TA y LTA presentan colina, un amino alcohol abundante en eucariotas y con escasa presencia en bacterias (García *et al.*, 1998). En el caso concreto de neumococo la pared se encuentra rodeada por una cápsula compuesta de polisacáridos (Austrian, 1981) hacia la cual van dirigidas las vacunas existentes. La cápsula protege al microorganismo de ser fagocitado por las células del sistema inmunitario (Yother, 2004). Se han descrito más de 90 tipos serológicos de cápsula diferentes en base a la composición de las mismas (Henrichsen, 1995).

En las bacterias Gram-negativas (por ejemplo *Escherichia coli* y *Vibrio cholerae*), el PG constituye alrededor del 10% de pared, estando constituido el resto por una membrana externa. La pared celular de estas bacterias consta de una fina capa de PG que se encuentra situada en el periplasma, flanqueado por la membrana citoplasmática interna y la membrana externa. Este tipo de bacterias no posee ácidos teicoicos en su PG; sin embargo, presenta una capa adicional compuesta de lipopolisacáridos (LPS).

Según un informe del Centro Americano de Control y Prevención de Enfermedades (CDC), las bacterias Gram-negativas son patógenos especialmente preocupantes ya que se han vuelto resistentes a casi todos los fármacos existentes para su tratamiento. Aunque la resistencia también ocurre en organismos Gram-positivos, el grado de resistencia sigue siendo menor que en Gram-negativos, según el CDC. Los patógenos Gram-negativos que causan mayores problemas son las especies *Enterobacteriaceae*, *Pseudomonas aeruginosa* y *Acinetobacter*. El tratamiento de infecciones causadas por estos organismos se ha convertido en un reto diario en los hospitales debido a las múltiples resistencias que han generado en poco tiempo.

## 1.2. BIOSÍNTESIS DE PEPTIDOGLICANO



**Figura 1.3. Esquema de la biosíntesis del peptidoglicano.** Los precursores del PG se sintetizan en el citoplasma, una vez formado el Lípido II es trasladado al espacio periplásmico a través de una flipasa. En el espacio periplásmico se produce la glicosilación y la transpeptidación mediada por las proteínas PBPs.

La ruta de biosíntesis del peptidoglicano se lleva a cabo en tres etapas principalmente, la síntesis del precursor UDP-MurNAc-pentapéptido en el citoplasma celular, la formación del lípido II y su traslocación hacia el espacio periplásmico y, por último, el crecimiento del peptidoglicano mediante el ensamblaje de las unidades anteriores (Fig. 1.3). En el proceso, que se inicia a partir del precursor uridina difosfato (UDP), se van formando las unidades de UDP-NAG y UDP-NAM (MurA cataliza este paso, en Gram negativas solo hay una copia de este gen y su depleción resulta letal) y a continuación se va adicionando L-Ala, D-Glu, L-Lys o meso-Diaminopimélico (DAP) y D-Ala. Todas estas reacciones son catalizadas por enzimas de la familia Mur.

Posteriormente, una proteína integral de membrana (MraY) se encarga de anclar el NAM-pentapéptido a la membrana citoplasmática a través del transportador undecaprenil-fosfato (UP), generándose lo que se conoce comúnmente como Lípido I. El último paso de la ruta de biosíntesis, que tiene lugar en el citoplasma, es la conversión del lípido I en lípido II a través de la transferencia de un UDP-NAG.



La transferencia del lípido II al exterior celular no se produce de manera espontánea, se cree que está mediada por proteínas implicadas en los procesos de división y elongación celular, la FtsW y la RodA (Pastoret *et al.*, 2004). Por otro lado, estas proteínas podrían estar actuando como flipasas, dirigiendo la exportación del lípido II hacia el espacio periplásmico.

Finalmente tiene lugar la polimerización del lípido II, la cual conlleva una etapa inicial en la que dos unidades de lípido II reaccionan para formar el lípido IV y una etapa de elongación, en la que se van adicionando sucesivamente unidades de lípido II. Las PBPs (del inglés *Penicillin Binding Proteins*) son las enzimas que participan en este proceso y se encargan de polimerizar las cadenas formando nuevos enlaces  $\beta$ -1,4 NAM-NAG. El entrecruzamiento de las cadenas glicánicas tiene lugar a través de la transpeptidación de las unidades peptídicas de las cadenas glicánicas adyacentes. Las enzimas transpeptidasas son una diana natural para los antibióticos  $\beta$ -lactámicos, como son las penicilinas y las cefalosporinas.

Las PBPs se pueden dividir en dos tipos: las HMM-PBPs (*High Molecular Mass PBPs*) y las LMM-PBPs (*Low Molecular Mass PBPs*). Las HMM-PBPs son proteínas que se encuentran ancladas a la membrana, poseen dos dominios, uno con actividad transglicosilasa y otro con actividad transpeptidasa, pudiendo realizar ambas acciones independientemente. El dominio transpeptidasa se encarga de entrecruzar distintas cadenas glicánicas a través de los péptidos del NAM. Las PBPs actúan como D,D-transpeptidasas, uniendo el grupo carboxilo de la D-Ala en la posición 4, y el grupo amino del m-DAP en la posición 3 de la otra cadena.

Las LMM-PBPs regulan el grado de transpeptidación, estas enzimas poseen un único dominio transpeptidasa (con actividad D-Ala-D-Ala carboxipeptidasa), y tienen como único sustrato el extremo D-Ala-D-Ala del pentapéptido. Su función es eliminar el último D-Ala para que no se pueda producir el enlace cruzado entre las dos cadenas glicánicas, ya que el tetrapéptido resultante no sirve como sustrato para las D-Ala-D-Ala transpeptidasas (HMM PBPs) (Sauvage *et al.*, 2008).

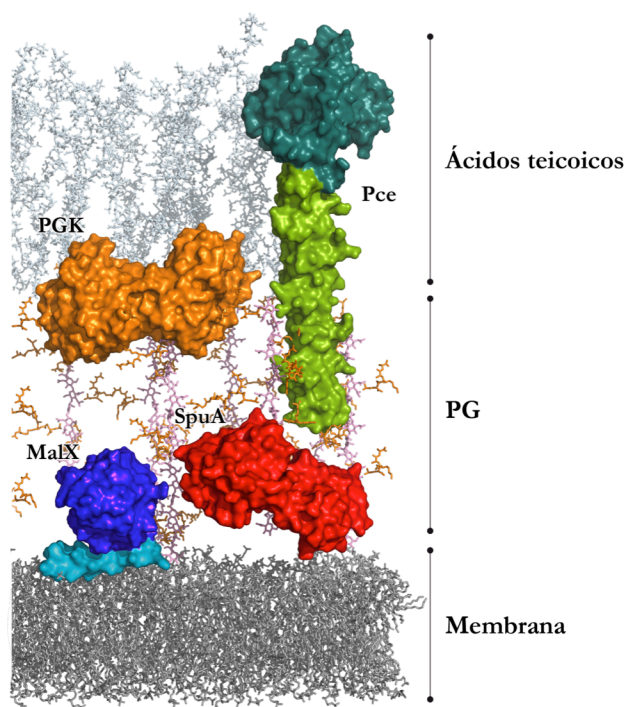
A pesar de que es común encontrar enlaces cruzados 4–3 entre cadenas adyacentes de PG, reacción catalizada por D,D-transpeptidasas, en ciertas cepas bacterianas se observan también enlaces 3–3, formados por enzimas L,D-transpeptidasas. El enlace prevalente en el PG de *E. coli* es el que forman las D,D-transpeptidasas, los enlaces minoritarios 3-3 se han detectado en la fase de crecimiento exponencial ( $\sim 2\%$ ) y estacionaria ( $\sim 4\%$ ) (Pisabarro *et al.*, 1985, Driehuis & Wouters, 1987, Blasco *et al.*, 1988). Se ha demostrado que en *Clostridium difficile* los enlaces cruzados formados por enzimas L,D-transpeptidasas aumentan en presencia de ampicilina, lo cual indica que este antibiótico no las inhibiría (Peltier *et al.*, 2011). Además del papel que juegan las L,D-transpeptidasas en la biosíntesis del PG, esta familia de proteínas es responsable de la unión de la lipoproteína Braun al PG de *E. coli* (Magnet *et al.*, 2007), la

proteína Braun es una de las más abundantes en las membranas de las bacterias Gram-negativas y es responsable de la integridad estructural de la membrana externa.

### 1.3. PROTEÍNAS DE SUPERFICIE

Las proteínas de superficie de bacterias patógenas como el neumococo son consideradas factores potenciales de virulencia debido a que participan activamente en las interacciones hospedador-patógeno.

Se han identificado tres grupos mayoritarios de proteínas presentes en la superficie del neumococo: entorno a 50 tipos de lipoproteínas, más de 18 proteínas con secuencia consenso LPXTG unidas covalentemente a través de una sortasa al PG, y más de 16 proteínas de unión a colina (CBPs) (Perez-Dorado *et al.*, 2012). Además de estos tres grupos principales, en la envoltura del neumococo podemos encontrar otro grupo de proteínas que no presenta un péptido líder ni un motivo de unión a membrana; se le ha denominado proteínas de superficie no clásicas (NCSP) (Fig. 1.4). En estudios recientes se ha demostrado que estas NCSP podrían jugar un papel muy importante en la virulencia del patógeno (Bergmann & Hammerschmidt, 2006, Perez-Dorado *et al.*, 2012).



**Figura 1.4. Proteínas de superficie presentes en neumococo.** Existen cuatro grupos de proteínas que se pueden encontrar en la pared del neumococo: lipoproteínas (por ej. MalX), proteínas LPXTG (por ej. SpuA), proteínas de unión a colina (por ej. Pce) y proteínas no-clásicas de membrana (por ej. PGK). Figura adaptada de (Perez-Dorado *et al.*, 2012).

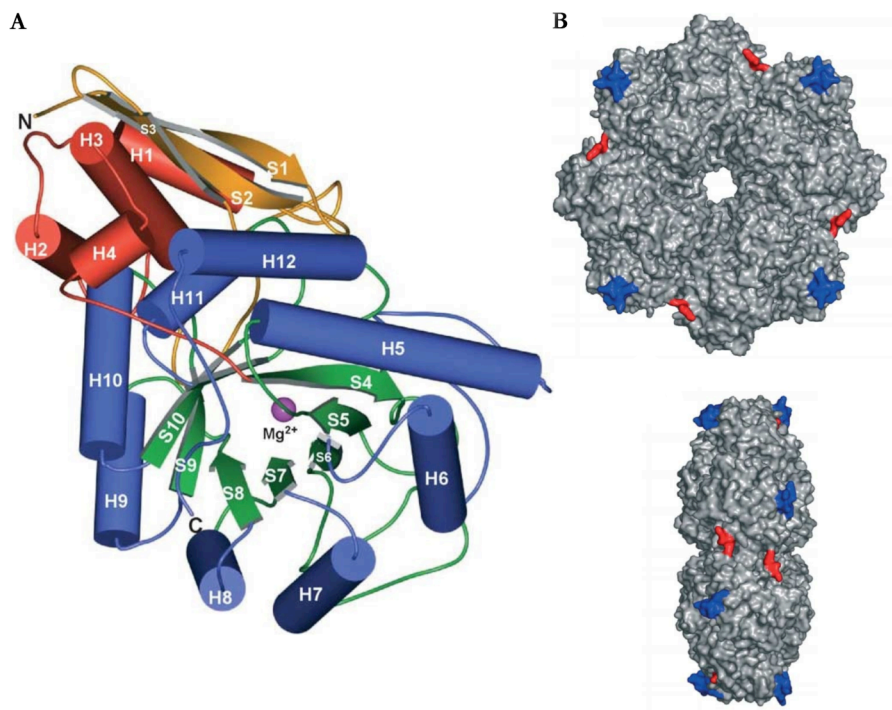
#### 1.3.1 UNA NUEVA FAMILIA: LAS PROTEÍNAS NCSP

Existe un conjunto de proteínas de superficie de neumococo que carecen de una señal de anclaje a membrana o una señal secretora. A estas proteínas se las conoce como proteínas no clásicas de superficie

(NCSP), todas ellas son proteínas citoplásmicas con papeles intracelulares y no se las relacionaba con la interacción hospedador-patógeno. Sin embargo, se ha demostrado que estas proteínas NCSP se encuentran en la superficie celular y muestran funciones diversas en interacción con el hospedador humano, frecuentemente actúan como adhesinas. Por este motivo se las conoce como *moonlighting proteins*, debido a su capacidad para realizar múltiples funciones.

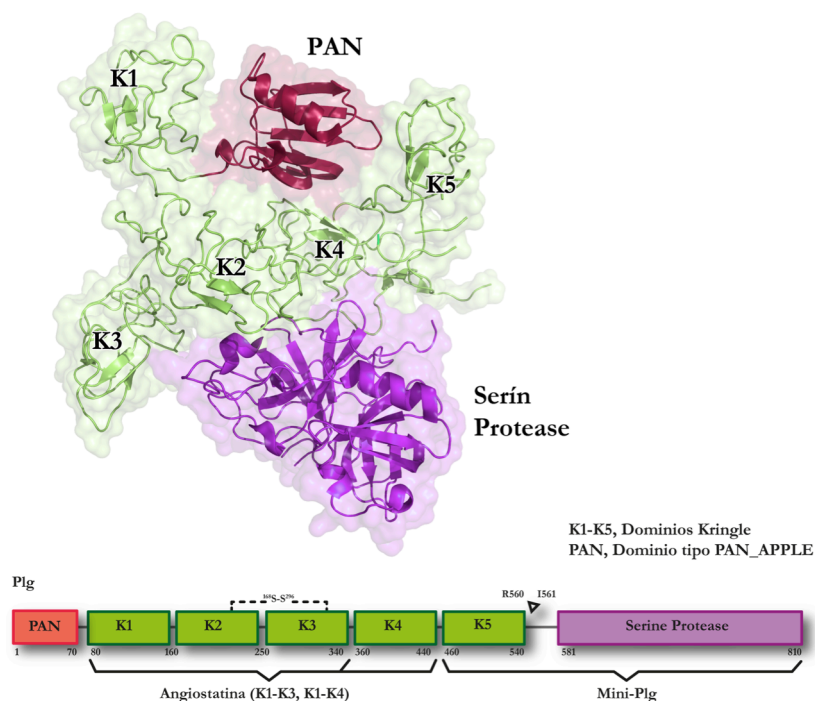
Las primeras proteínas capaces de realizar múltiples funciones fueron las cristalinas (proteínas estructurales que están presentes en las lentes de ojos de vertebrados) descritas en 1989 por Piatigorsky y Wistow (Piatigorsky & Wistow, 1989). Las proteínas multifuncionales son capaces de realizar tareas de manera autónoma, a veces no relacionada, sin necesidad de diversificar las funciones en dominios. Existen proteínas multifuncionales en muchas especies, incluyendo plantas, animales, levaduras y procariotas. En *Mycobacterium tuberculosis* se encuentra la glutamato racemasa (MurI) que, además de llevar a cabo un papel importante en la biosíntesis del peptidoglicano a través de la conversión del ácido L-glutámico en D-glutámico, actúa como inhibidor de la ADN girasa (Sengupta *et al.*, 2008).

El neumococo presenta diferentes moléculas en su superficie que interaccionan con receptores del hospedador, proteínas de la matriz extracelular y compuestos del sistema circulatorio (Bergmann & Hammerschmidt, 2006, Perez-Dorado *et al.*, 2012). La interacción de las proteínas NCSP con moléculas del hospedador promueve la invasión del patógeno y la dispersión de la infección, como es el caso de una serie de enzimas metabólicas bacterianas como la enolasa y la gliceraldehído-3-fosfato deshidrogenasa (GAPDH) que funcionan como NCSP con papeles diversos (Hughes *et al.*, 2002): como la adherencia a diversos componentes del hospedador, como la fibronectina (Donofrio *et al.*, 2009) y la laminina (Carneiro *et al.*, 2004), así como a receptores de PLG (Bergmann & Hammerschmidt, 2007, Bergmann *et al.*, 2001) (Fig. 1.5).



**Figura 1.5. Estructura tridimensional de la enzima enolasa de neumococo.** a) Monómero que compone la estructura completa de la enolasa. Se representan las hélices como cilindros y el ión magnesio como esfera morada. b) Representación de la superficie del octámero de enolasa con los sitios de unión a plasminógeno marcados en azul y rojo. En la parte superior se muestra una vista frontal y en la inferior una lateral.

En los últimos diez años se ha investigado intensamente la activación y unión a la proteína más importante de la fibrinólisis, el plasminógeno (PLG). Su reclutamiento en la superficie del neumococo provoca un aumento en la actividad proteolítica de la serín proteasa plasmina, favoreciendo la dispersión intercelular y diseminación del patógeno a través de los tejidos y del sistema vascular (Bergmann & Hammerschmidt, 2007, Bergmann *et al.*, 2005). El PLG (92 kDa) es el precursor inactivo de la proteasa de amplio espectro plasmina, la cual degrada muchas de las proteínas presentes en el plasma sanguíneo. Está compuesto de un péptido de pre-activación (~8 kDa), cinco estructuras denominadas *kringle* (K1-5, 65 kDa) y un dominio serín proteasa (Miyashita *et al.*, 1988). Los cuatro primeros *kringles* conforman la angiostatina (Chavakis *et al.*, 2005). El último *kringle* junto con la plasmina se denomina mini plasminógeno (mini-PLG) y es una molécula inhibidora de la angiogénesis, (Duboscq *et al.*, 1997) (Fig. 1.6). El PLG se convierte en plasmina a través de los activadores de plasminógeno, uPA (*urokinase-type plasminogen activator*) y tPA (*tissue-type plasminogen activator*) (Castellino & Powell, 1981), la cual puede hidrolizar las cadenas de fibrina. Además, se ha detectado un incremento en los niveles de uPA y tPA en el fluido cerebroespinal en pacientes que sufren de meningitis (Winkler *et al.*, 2002).

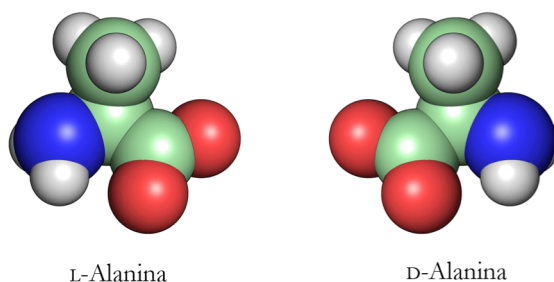


**Figura 1.6. Estructura tridimensional del PLG humano** (código PDB 4DUR). Elementos que componen el PLG: el péptido de activación se muestra en rojo, los dominios *kringle* en verde y el dominio de la proteasa en morado.

El rol multifuncional que presentan muchas de las proteínas NCSP relacionado con virulencia sigue sin conocerse; sin embargo, los análisis estructurales de las mismas podrán ayudar a comprender mejor su implicación en virulencia convirtiéndose éstas en nuevas dianas para el desarrollo de vacunas y/o nuevos antibióticos.

#### 1.4. RELEVANCIA DE LOS D-AMINOÁCIDOS EN LA FISIOLÓGÍA BACTERIANA

Una molécula quiral es aquella que no puede superponerse a su reflexión especular. La palabra quiral proviene del griego *cheir* ( $\chi\epsilon\iota\rho$ ), que significa “mano”. Se dice que los objetos quirales (incluyendo las moléculas) se parecen a las “manos” (Fig. 1.7). El término quiral se utiliza para describir a las moléculas porque los enantiómeros se relacionan entre sí en la misma forma que la mano derecha se relaciona con la izquierda. La quiralidad está ampliamente distribuida a nivel molecular, en 1848 Louis Pasteur descubrió la quiralidad molecular en cristales de ácido tartárico (Flack, 2009). Nueve años después, mientras estudiaba la fermentación del ácido tartárico por microorganismos, Pasteur descubrió que una de las imágenes especulares del ácido tartárico se consumía con mayor preferencia que la otra (Gal, 2008). En años sucesivos la gran importancia de la quiralidad en los procesos bioquímicos se fue clarificando.



**Figura 1.7. Moléculas quirales.** Imagen especular que muestra el aminoácido alanina en su conformación L y D.

Las dos imágenes especulares de una molécula quiral se denominan isómeros o enantiómeros (del griego *εχθρός*, enemigo). La mayoría de las propiedades físicas de los enantiómeros son idénticas; sin embargo, tienen diferentes interacciones con los planos de luz polarizada. Cuando la luz polarizada pasa a través de una solución con moléculas quirales, unos enantiómeros rotarán el plano de luz polarizada en sentido de las agujas del reloj (mano derecha), mientras que otros enantiómeros la rotarán en el sentido contrario (mano izquierda) (Fig. 1.7). En las moléculas, la representación básica más común de quiralidad es un átomo de carbono con cuatro sustituyentes diferentes (carbono asimétrico), este carbono es el centro quiral. La disposición tridimensional de los átomos unidos al centro quiral se emplea para describir la quiralidad de la molécula. Existen diferentes nomenclaturas para describir la quiralidad de las moléculas, la más empleada para designar aminoácidos y azúcares es la que los clasifica como D y L (Meierhenrich, 2008). Mientras que por síntesis química se producen ambos enantiómeros, muchas enzimas presentan una marcada selectividad de sustrato y, como consecuencia, muchos procesos bioquímicos utilizan y generan enantiómeros particulares. Por ejemplo, los L-aminoácidos son los constituyentes naturales de las proteínas, sin embargo, los D-aminoácidos (DAA) no se pueden incorporar a las proteínas mediante síntesis en los ribosomas.

Los DAA juegan un papel importante en la señalización intercelular; se han encontrado altas concentraciones de D-aspartato y de D-serina en el sistema nervioso central de mamíferos y humanos (Dunlop *et al.*, 1986, Fisher *et al.*, 1991, Hashimoto *et al.*, 1993). La D-serina actúa como un co-agonista de los receptores de glutamato en el cerebro, los cuales están relacionados con el aprendizaje, la memoria y el comportamiento en mamíferos (Kleckner & Dingledine, 1988, Wolosker, 2007, Wolosker *et al.*, 2002). Asimismo, el metabolismo de la D-serina está asociado con diferentes desordenes neurodegenerativos como la esquizofrenia, la isquemia y la epilepsia (Fuchs *et al.*, 2005).

En la mayoría de las bacterias los únicos D-aminoácidos que se producen son D-Ala y D-Glu, los cuales se incorporan al PG (Holtje, 1998). En *Vibrio cholerae*, una racemasa específica, BsrV, produce D-

Met y D-Leu, también *Bacillus subtilis* presenta otra racemasa que produce D-Tyr y D-Phe; estos D-aminoácidos no canónicos (NCDAA, del inglés *Non-canonical D-aminoacids*) parece que tienen un papel regulador en la síntesis de PG. Los D-aminoácidos influyen en la composición, cantidad y fortaleza del PG, bien a través de su incorporación directa en el PG o a través de la regulación de las enzimas que intervienen en su síntesis y modificación. Debido a que la acumulación de los D-aminoácidos coincide con la transición a la fase estacionaria y tiene como función des-regular la síntesis de PG, los D-aminoácidos podrían activar la ralentización del metabolismo en la pared celular y en los compartimentos celulares cuando los recursos son escasos. Por ello, la síntesis de D-aminoácidos podría ser una estrategia de las bacterias para adaptarse a las condiciones cambiantes del medio. (Lam *et al.*, 2009).

*E. coli* es capaz de incorporar NCDAA en su PG a pesar de que esta bacteria no produzca o libere NCDAA, así los D-aminoácidos podrían actuar sobre la fisiología del PG de otras especies bacterianas diferentes a las que los han producido (Cava *et al.*, 2011).

Los DAA están directamente involucrados en gran variedad de procesos biológicos que tienen lugar en la bacteria, como es el crecimiento, la germinación de las esporas y la formación de biofilms. Prácticamente todas las bacterias forman biofilms para protegerse de las ofensas ambientales; los DAA pueden actuar como sensores evitando la formación de biofilms en *S. aureus* y *P. aeruginosa*, o promoviendo la disgregación de biofilms preformados en *B. subtilis* (Kolodkin-Gal *et al.*, 2010, Yang *et al.*, 2015). Los DAA también están involucrados en la esporulación de la bacteria, en especies de *Bacillus* se ha comprobado que el aminoácido D-Ala inhibe la germinación de esporas cuando hay una densidad de población alta (Lam *et al.*, 2009). Además, algunos patógenos están rodeados de una cápsula que les protege de la fagocitosis aumentando su virulencia, es el caso de la cápsula de *B. anthracis* que está compuesta de un polímero infrecuente, el ácido poli- $\gamma$ -D-glutámico ( $\gamma$ DPGA), producido después de la germinación de la espora (Hubbard *et al.*, 2013).

Podemos encontrar DAA en su forma libre así como contenidos en péptidos (Bodanszky & Perlman, 1969). La existencia de residuos con configuración D en los péptidos confiere resistencia frente al ataque de proteasas, las cuales presentan especificidad por los L-aminoácidos. A estos péptidos compuestos de DAA se les conoce como péptidos no ribosomales (NRP) y son sintetizados por las NRP sintetasas (NRPs), la síntesis de estos péptidos se desarrolla de manera alternativa al mecanismo tradicional de traducción. Los NRP están compuestos de estructuras cíclicas y/o ramificadas pudiendo contener D-aminoácidos, presentan radicales N-formilo y N-metilo, así como grupos glicosilados, halogenados, acilados o hidroxilados. Estos péptidos forman una familia diversa de productos naturales con actividades biológicas y propiedades farmacológicas de amplio espectro; sirviendo como ejemplo la actinomicina D que se emplea de manera comercial como antibiótico.



Por otro lado, las bacterias del suelo producen NRP activamente, es el caso de *Serratia plymuthica* una bacteria presente en suelo y en el rizoma de plantas de trigo, guisante y melón, entre otras, capaz de matar al nematodo *Caenorhabditis elegans* a través de la producción de metabolitos secundarios como la zeamina (Hellberg *et al.*, 2015). La importancia en la producción de este metabolito por parte de la bacteria recae en la dificultad para controlar las plagas mediante el uso de pesticidas tradicionales, convirtiendo a la bacteria en un potencial agente de control biológico.

Como se ha descrito anteriormente, los DAA son necesarios para el crecimiento de la bacteria y, además, son componentes esenciales de la pared celular bacteriana pero también son determinantes en la regulación de procesos biológicos importantes para la supervivencia bacteriana como son el desensamblaje del biofilm y la esporulación.

#### **1.4.1. DESENSAMBLAJE DEL BIOFILM**

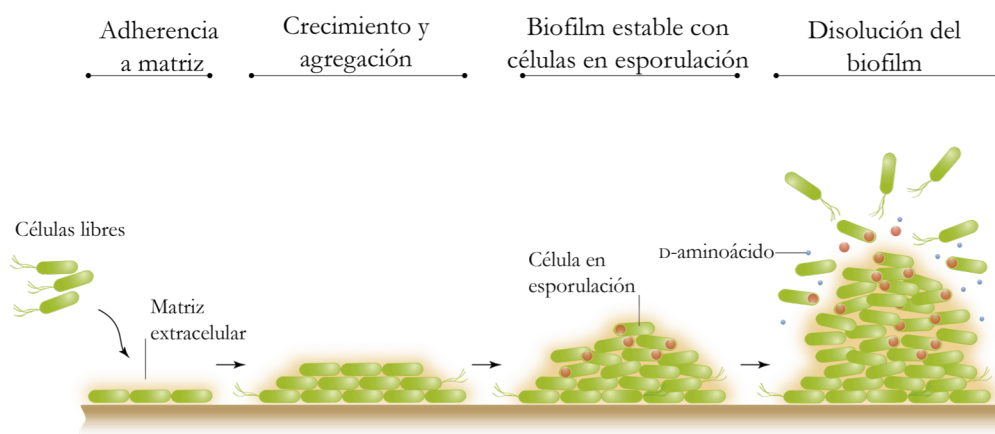
Las bacterias forman comunidades multicelulares conocidas como biofilms que, con el tiempo, acaban disgregándose. La formación de estos biofilms permite a las bacterias protegerse de las agresiones ambientales (Karatan & Watnick, 2009). Sin embargo, a medida que el biofilm crece los nutrientes se vuelven limitantes, los productos de desecho se acumulan y las bacterias necesitan recuperar el modo de vida libre, independiente de la comunidad del biofilm. Un claro ejemplo de organismo que adopta una forma de vida en comunidad es *Bacillus subtilis* estableciéndose en superficies semisólidas y formando finas películas en la interfase aire-líquido de los cultivos verticales (Branda *et al.*, 2001, Vlamakis *et al.*, 2008, Aguilar *et al.*, 2007). Las células en el biofilm se mantienen juntas gracias a una matriz extracelular consistente en exopolisacárido y fibras amiloides (Branda *et al.*, 2006, Branda *et al.*, 2001, Romero *et al.*, 2010). Los biofilms están presentes en ambientes naturales pero también en la industria y en los hospitales (Liu & Post, 2009, Kassir *et al.*, 2009, Homoe *et al.*, 2009, Jacobsen *et al.*, 2008, Kobayashi, 2005, Reslinski *et al.*, 2009, Yankah *et al.*, 2005). *Acinetobacter baumannii* es un organismo capaz de sobrevivir en hospitales a pesar de encontrarse en un ambiente altamente estéril, tanto la supervivencia en estos ambientes como la virulencia del patógeno están asociadas con su capacidad para formar biofilms (Rodríguez-Bano *et al.*, 2008). Las infecciones bacterianas asociadas a los biofilms adheridos a los aparatos médicos, como son los catéteres intravenosos, son muy difíciles de erradicar ya que las matrices extracelulares protegen a las bacterias formadoras del biofilm de los antimicrobianos y la respuesta inmune. Por esta razón, se requiere del diseño de nuevos agentes que prevengan la formación de biofilms y/o desestabilicen los ya formados.

Recientemente se ha demostrado que la disolución del biofilm es inducida por una mezcla de D-aminoácidos (D-Leu, D-Met, D-Trp y D-Tyr) producidos por *B. subtilis* dentro del biofilm (Kolodkin-Gal *et al.*, 2010). Además, se ha observado que la acumulación de estos D-aminoácidos se produce únicamente



en biofilms maduros (6-8 días). Sin embargo, si se añaden DAA exógenos a cultivos celulares de *B. Subtilis* se previene la formación de biofilm.

Los D-aminoácidos actuarían como moléculas señal que promoverían la disgregación del biofilm en *B. Subtilis* cuando la comunidad bacteriana se encontrara en condiciones de escasez de nutrientes y acumulación de productos de desecho, por lo que, recuperar la forma de vida libre resultaría beneficioso (Fig. 1.8).

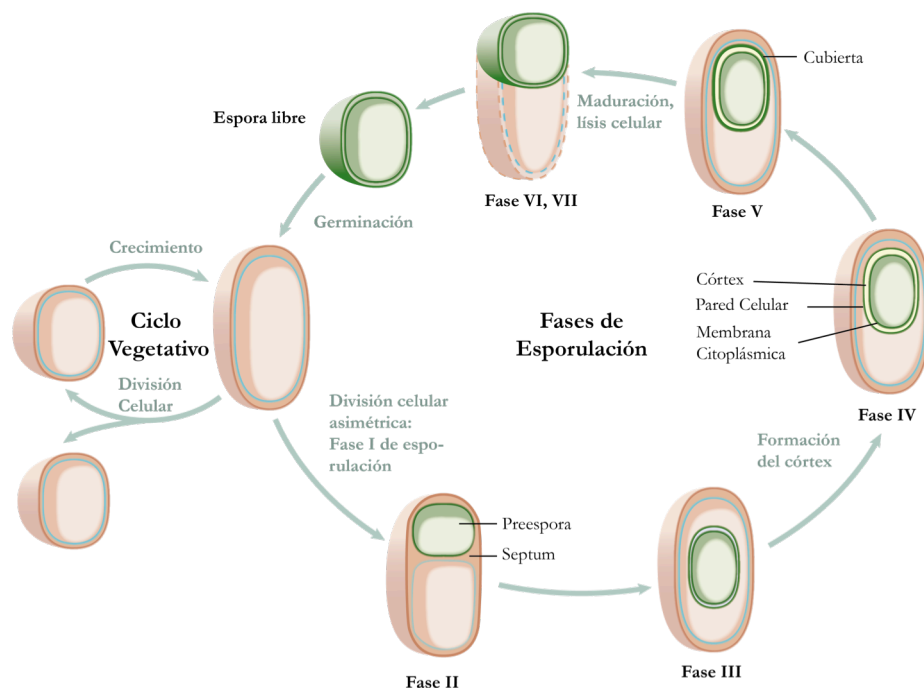


**Figura 1.8. Ensamblaje y disolución del biofilm.** La formación del biofilm tiene lugar en varias etapas, desde el crecimiento y maduración hasta el desensamblaje de la comunidad bacteriana. Inicialmente las células libres con flagelo se inmovilizan sobre una superficie produciendo matriz extracelular. En los biofilms maduros las células de la matriz producen esporas. Con el tiempo las células del biofilm producen D-aminoácidos que rompen la matriz formada permitiendo la dispersión de las células en el ambiente. (Adaptado de (Vlamakis *et al.*, 2013)).

El mecanismo mediante el cual se produce la dispersión del biofilm se desconoce por ahora. Se cree que los D-aminoácidos podrían inducir la disgregación de las fibras amiloides asociadas a la matriz que mantiene las células de *B. subtilis* en el biofilm (Kolodkin-Gal *et al.*, 2010). Sin embargo, el uso de D-aminoácidos para combatir infecciones asociadas a la formación de biofilms podría ser una estrategia prometedora, ya que los D-aminoácidos parecen tener propiedades farmacocinéticas y carecen de toxicidad (Jayaraman & Wood, 2008).

#### 1.4.2. EFECTO DE LOS D-AMINOÁCIDOS EN ESPORULACIÓN

El proceso de esporulación en bacterias es una respuesta adaptativa al estrés ambiental, como puede ser la falta de nutrientes, el cual se desarrolla de manera controlada y programada. La esporulación comienza con una división celular asimétrica fruto de la cual se genera una endoespora que deviene en espora, esta espora obtiene nutrientes de la célula madre la cual termina lisándose a través de la muerte celular programada (Stragier & Losick, 1996) (Fig. 1.9).

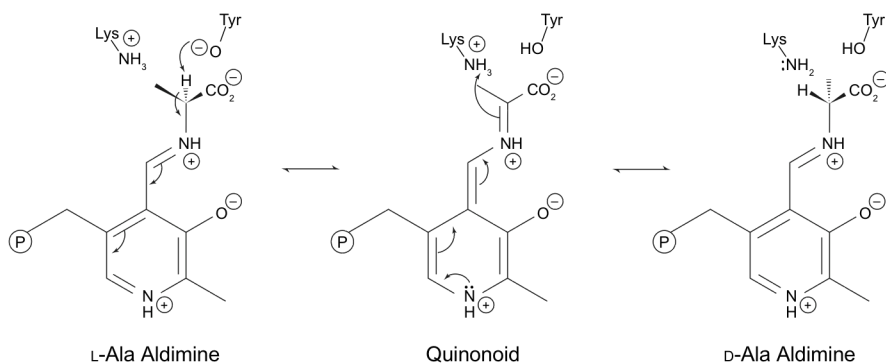


**Figura 1.9. Ciclo de esporulación.** Esquema donde se presentan las diferentes fases por las que pasa una bacteria durante el proceso de esporulación.

Las esporas bacterianas se encuentran en un estado metabólico latente y son capaces de resistir condiciones extremas, incluyendo las altas temperaturas, la radiación, la desecación, el pH extremo y las toxinas, entre otros (Setlow, 2003). La espora está rodeada de una cubierta que actúa como barrera frente a moléculas tóxicas como la lisozima; además, el 20% de peso seco de la endoespora es dipicolinato de calcio, el ácido dipicolínico podría ser el responsable de la resistencia a altas temperaturas y el calcio podría ayudar en la resistencia al calor y a los agentes oxidantes (Driks, 1999). En presencia de promotores específicos como la L-Ala u otros nutrientes, las esporas pueden reactivar su metabolismo y crecer en estado vegetativo. En 1949 se observó que la D-Ala era un potente inhibidor de la germinación de esporas en muchas especies de *B. subtilis* (Hills, 1949). En trabajos sucesivos se demostró que las especies de *Bacillus* utilizan D-Ala como auto inhibidor de la germinación de esporas cuando hay muchas esporas. Esta actividad está mediada por la expresión de una alanina racemasa que convierte el agente promotor (L-Ala) en agente inhibidor (D-Ala) de la formación de esporas (Halvorson & Spiegelman, 1952). Este mecanismo de auto inhibición podría ser un mecanismo adaptativo para prevenir la germinación prematura bajo condiciones de escasez de nutrientes y alta densidad de población, una condición ambiental que causaría la muerte celular. En un trabajo más reciente se sugiere que la D-Ala podría, además, alterar la cinética de germinación *in vivo* para favorecer la eficiencia y el tiempo de infección; en *Bacillus anthracis* la coordinación del tiempo de germinación es importante durante las primeras etapas de infección y puede determinar la supervivencia de la célula o sucumbir al ataque de los macrófagos del hospedador (McKevitt *et al.*, 2007).

### 1.5. LAS RACEMASAS, ENZIMAS IMPLICADAS EN LA PRODUCCIÓN DE DAA

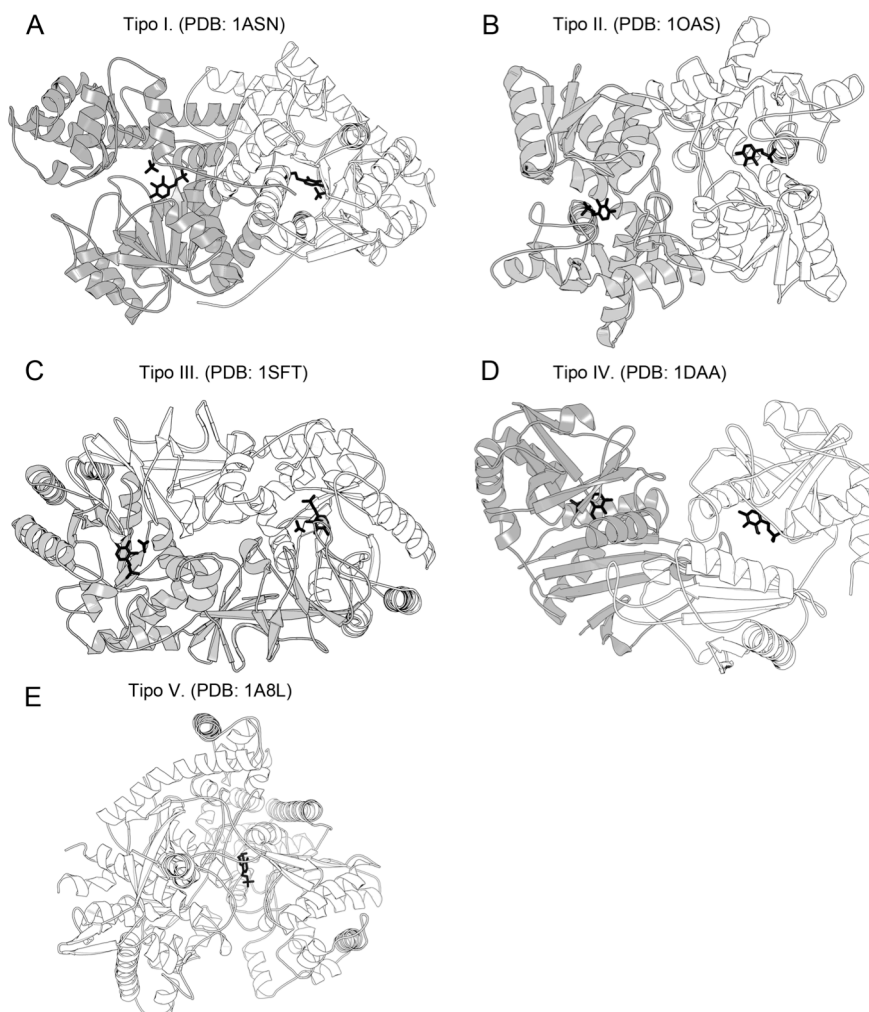
Las racemasas son las enzimas que catalizan la conversión de los aminoácidos L en D. Estas enzimas se pueden clasificar en dos grupos principales en función de si requieren del coenzima piridoxal 5'-fosfato (PLP), la forma activa de la vitamina B6, para realizar la catálisis. Dentro del grupo de las racemasas PLP dependientes están contenidas enzimas eucariotas como la aspartato racemasa (AspR), alanina racemasa (AlaR) y la serina racemasa (SerR); así como las procariotas AlaR, SerR, ArgR y las recientemente identificadas como una nueva clase de racemasas de amplio espectro (BsrR). Dentro de las racemasas PLP-independientes de procariotas se encuentran las aspartato racemasas (AspR), glutamato racemasas (GluR) y las prolina racemasas (ProR). Ambos tipos de enzimas, PLP-dependientes y PLP-independientes, desprotonan el carbono  $\alpha$  del L-aminoácido para re-protonarlo en el lado contrario generando la forma D del aminoácido.



**Figura 1.10. Mecanismo de reacción de las racemasas.** La reacción comienza con la formación de una aldimina mediante un mecanismo de transaldiminación de la lisina unida al PLP. La AlaR usa la tirosina para desprotonar la L-Ala formando una quinona como intermediario de reacción, el cual es re-protonado en el lado opuesto al cofactor a través de la lisina para producir la D-Ala.

El mecanismo de reacción de la racemasa AlaR es uno de los más estudiados, en este caso se forma una base de *Schiff* interna a través de un enlace covalente entre el PLP y un residuo de Lys de la enzima. La L-Ala reacciona con el PLP para formar una base de *Schiff* externa mediante una reacción de transaldiminación. La extracción del hidrógeno  $\alpha$  del sustrato produce una forma aniónica estabilizada como intermediario quinona con el PLP (Fig. 1.10). A través de una segunda reacción de transaldiminación con la Lys el intermediario quinona es entonces protonado liberando la forma opuesta del residuo de Ala (Eliot & Kirsch, 2004).

Se han descrito cinco tipos de enzimas PLP-dependientes atendiendo al tipo de plegamiento, tipo I-V (Grishin *et al.*, 1995, Soda *et al.*, 2001). Las racemasas se encuentran contenidas dentro del tipo III, que contiene AlaR, ArgR y alguna aminoácido descarboxilasa (Fig. 1.11), siendo la enzima AlaR el principal representante. AlaR actúa como homo-dímero, cada monómero se compone de dos dominios, uno de ellos con un plegamiento en forma de barril  $\alpha/\beta$  de ocho hebras y el otro contiene hebras  $\beta$ . El sitio activo del dímero está formado por residuos de ambos monómeros.



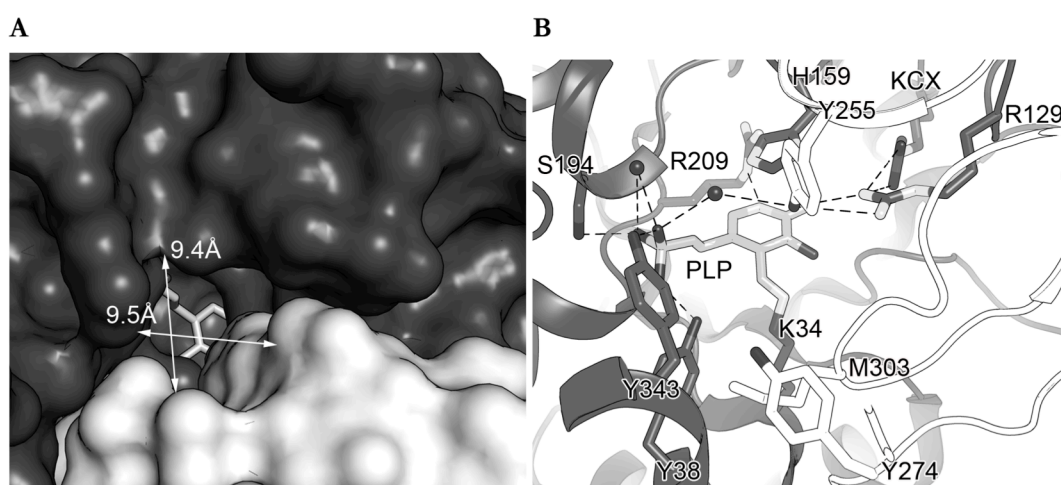
**Figura 1.11. Tipos de plegamiento de las enzimas PLP-dependientes (Tipos I-V).** Las estructuras de estas enzimas se pueden relacionar con el tipo de reacción que llevan a cabo. El coenzima PLP está representado en modelo de palos negros.

### 1.5.1. ALANINA RACEMASAS

Las alanina racemasas (AlaR) pertenecen al tipo III (Fig. 1.11 c) de las racemasas PLP-dependientes. Se ha estudiado activamente el mecanismo de acción de las AlaR ya que son las responsables de la producción de la D-Ala, uno de los principales componentes de las cadenas peptídicas que conforman la

pared celular de bacterias Gram-positivas y Gram-negativas, además son importantes dianas para nuevas drogas antibacterianas.

La enzima actúa como homodímero, cada monómero se compone de dos dominios diferentes. El dominio N-terminal se pliega en forma de barril  $\alpha/\beta$  de ocho hebras y el dominio C-terminal presenta seis hebras  $\beta$  en forma de barril y dos hebras  $\beta$  adicionales (una de ellas del N-terminal y la otra del C-terminal) extendiendo así la superficie de este dominio. El sitio activo se encuentra contenido en el dominio N-terminal y está parcialmente cubierto por el dominio C-terminal del otro monómero, la cavidad resultante es estrecha y solo permite la entrada del sustrato L-Ala (Fig. 1.12 a)



**Figura 1.12. Sitio activo de las alanina racemasas.** a) Superficie molecular de la entrada al sitio catalítico de AlaR (AlaR<sub>EC</sub>) de *E. coli*, los monómeros están coloreados en blanco y negro. Se indican las dimensiones de la entrada de la cavidad en Armstrongs. b) Detalle del sitio activo de AlaR<sub>EC</sub>. Los residuos implicados en la catálisis, así como el cofactor están representados en modelo de varillas, las moléculas de agua se representan en esferas.

*Sitio activo de AlaR.* El cofactor PLP se encuentra fuertemente estabilizado en el sitio activo. Un residuo de lisina (Lys 34 en *Escherichia coli* AlaR, AlaR<sub>EC</sub>) forma un enlace covalente con el PLP (Fig 1.12 b). El anillo del PLP está flanqueado por residuos hidrofóbicos de un lado y por una histidina (His159 en AlaR<sub>EC</sub>) del otro (Fig 1.12 b). Un residuo de arginina (Arg129 en AlaR<sub>EC</sub>) se orienta hacia el PLP jugando un papel importante en la estabilización del carboxilato de la L-Ala en el intermediario quinona que se forma durante la reacción. Además en el sitio activo se encuentra un residuo peculiar, una lisina N-carboxilada (KCX), la cual establece un enlace de hidrógeno con la Arg129.

Mediante este trabajo de investigación se ha llevado a cabo la cristalización y resolución estructural de representantes de dos familias de proteínas de superficie presentes en bacterias: racemasas PLP dependientes y proteínas NSCP. Los objetivos propuestos comprenden la caracterización tanto de los sitios activos de las enzimas como de los sitios de interacción con otras proteínas, así como la

identificación del mecanismo de reacción y de interacción. Se ha realizado un análisis comparativo de las diferentes racemasas PLP conocidas con el fin de describir el mecanismo catalítico que define a las enzimas monoespecíficas y multiespecíficas. Además, a través de un análisis bioinformático, se ha comprobado la plasticidad de la racemasa de amplio espectro BsrV. Los resultados obtenidos se encuentran reflejados en los trabajos de investigación publicados, así como en el material inédito que se incluye en el apartado de Anexos.

## 2. OBJETIVOS





Con la finalidad de poder ampliar el conocimiento sobre la regulación de la biosíntesis de la pared bacteriana y los mecanismos de interacción de organismos patógenos con su hospedador, se han elegido varias proteínas implicadas en diferentes aspectos relacionados con la biosíntesis del peptidoglicano y la virulencia. En primer lugar, la fosfoglicerato quinasa (PGK), proteína citosólica que cataliza la segunda etapa de la glicólisis y que está implicada en la interacción con proteínas del hospedador. En segundo lugar, las racemasas de amplio espectro de diferentes microorganismos (BsrV, Bsu17640, BsrAb, BsrKo), enzimas capaces de racemizar una amplia variedad de aminoácidos, produciendo lo que se conoce como D-aminoácidos no canónicos que se incorporan al peptidoglicano de la bacteria. Por último, se ha estudiado una L,D-transpeptidasas de *E. coli* (YnhG) que actúa incorporando los D-aminoácidos a la pared bacteriana.

A continuación se enumeran los objetivos propuestos en este trabajo para cada una de las proteínas anteriormente descritas:

▪ **PGK de *Streptococcus pneumoniae*:**

- Cristalización y determinación estructural de la proteína PGK.
- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.
- Caracterización de la PGK como una proteína no clásica de superficie del neumococo y conocer su localización en la superficie del mismo.
- Caracterización de la interacción de PGK con proteínas del hospedador (plasminógeno) mediante una aproximación estructural y su implicación en virulencia.

▪ **BsrV de *Vibrio cholerae*:**

- Cristalización y determinación estructural de la proteína BsrV.
- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.
- Caracterización del mecanismo de reacción de las racemasas de amplio espectro y comparación con el de las alanina racemasas.
- Caracterización de su implicación funcional en el metabolismo de la bacteria.

▪ **Bsu1760 de *Bacillus subtilis*:**

- Cristalización y determinación estructural de la proteína Bsu17640.
- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.

▪ **BsrAb de *Acinetobacter baumannii*:**

- Cristalización y determinación estructural de la proteína BsrAb.

- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.

▪ **BsrKo de *Kingella oralis*:**

- Cristalización y determinación estructural de la proteína BsrKo.
- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.

▪ **YnhG de *Escherichia coli*:**

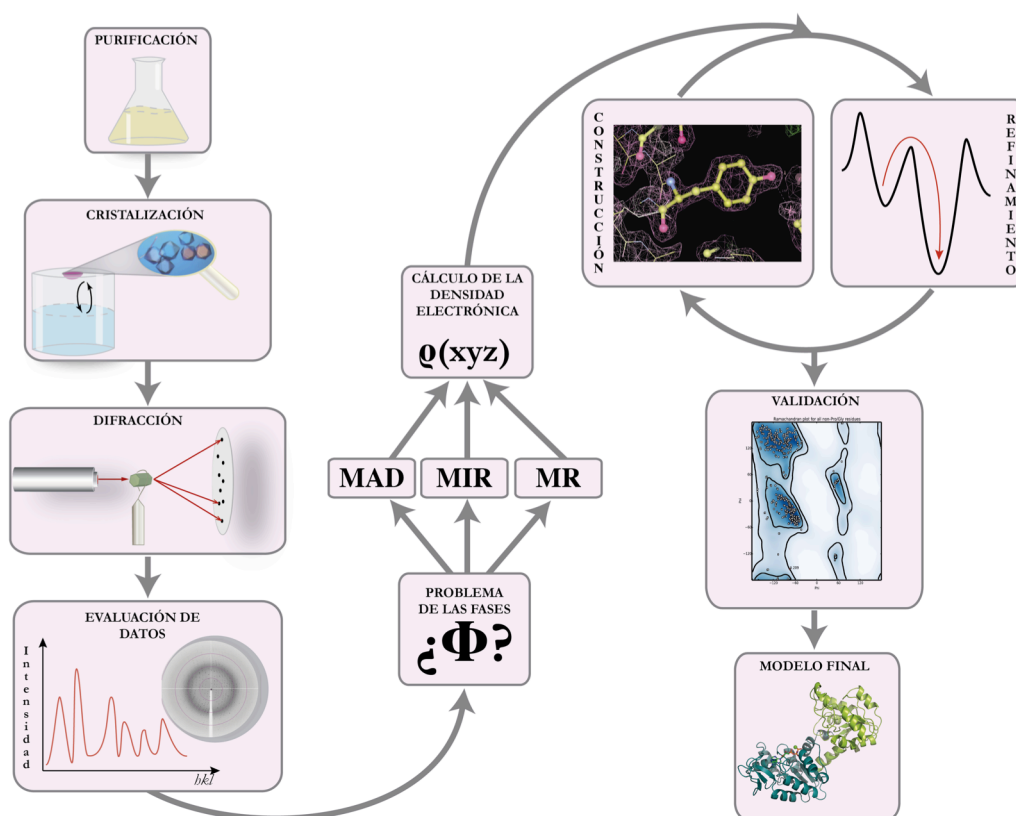
- Cristalización y determinación estructural de la proteína YnhG.
- Análisis estructural de dicha proteína, en su forma aislada y en presencia de ligandos de interés.

### 3. TÉCNICAS CRISTALOGRÁFICAS



### 3.1. VISIÓN GENERAL DEL PROCESO

La Cristalografía de Rayos X permite determinar la estructura tridimensional de la materia ordenada, tanto de naturaleza inorgánica como orgánica. Nos permite obtener una imagen tridimensional a escala atómica del material cristalizado. Aplicado al campo de proteínas, proporciona toda la información estructural, como el plegamiento y detalles físico-químicos como son las distancias y ángulos de enlace; dando una información clave para entender la función biológica y los procesos en los que están involucradas. Mediante la cristalografía se pueden estudiar desde sistemas sencillos hasta los más complejos, como la estructura de un virus o los grandes agregados moleculares, como es el caso del ribosoma. La determinación estructural de proteínas mediante Cristalografía de Rayos X conlleva una serie de etapas, las cuales se resumen a continuación (Fig. 3.1).



**Figura 3.1. Esquema general del proceso de resolución de estructuras moleculares y cristalinas mediante la difracción de rayos X.** Para la resolución estructural de proteínas es necesaria una muestra de proteína pura, una vez que se han obtenido cristales de proteína de alta calidad se miden en un difractómetro de rayos X o instalación sincrotrón, posteriormente se lleva a cabo la evaluación de los datos de difracción y tras varias etapas de refinamiento se obtiene un modelo final.

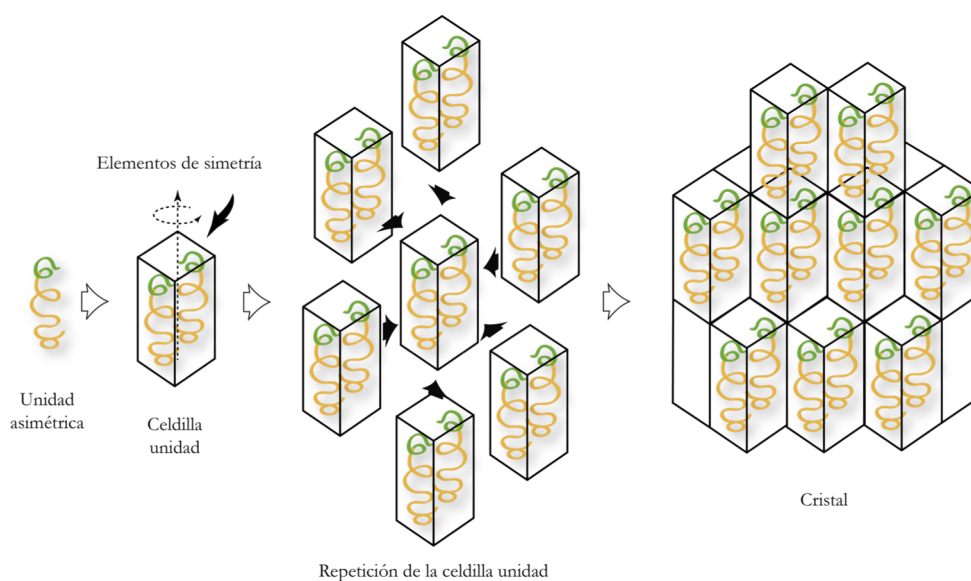
### 3.2. CRISTALIZACIÓN

La obtención de cristales es un paso crucial para la determinación de la estructura tridimensional. En el caso de las proteínas, este paso suele ser complejo debido a que las macromoléculas pueden contener zonas móviles, difíciles de estabilizar o incluso, que el orden alcanzado no sea suficiente para

dar lugar a un patrón de difracción de calidad suficiente. Los cristales de proteína contienen grandes cantidades de solvente (entre 30 y 80% del volumen del cristal (Matthews, 1968), lo que les confiere una extrema fragilidad y, en general, proporcionan un patrón de difracción débil. Sin embargo, el hecho de encontrar la proteína rodeada de solvente permite obtener su estructura en un ambiente no muy diferente al de su estado fisiológico. El objetivo de la cristalización es obtener cristales de tamaño y orden interno adecuados que proporcionen un espectro de difracción suficientemente completo como para determinar la estructura de las moléculas que lo componen.

### 3.2.1. CRISTAL Y SIMETRÍA

Un cristal es una red imaginaria formada por unidades elementales idénticas denominada cada una de ellas celdilla unidad. Estas unidades elementales se repetirán, por traslación, a lo largo de las tres direcciones del espacio generando así el volumen total del cristal. A su vez, la celdilla unidad está formada por una unidad mínima denominada unidad asimétrica, el contenido de la celdilla es el resultado de aplicar las operaciones de los *elementos de simetría* sobre la unidad asimétrica (Fig. 3.2). La celdilla unidad está definida por tres ejes ( $a$ ,  $b$ ,  $c$ ) y tres ángulos ( $\alpha$ ,  $\beta$ ,  $\gamma$ ).



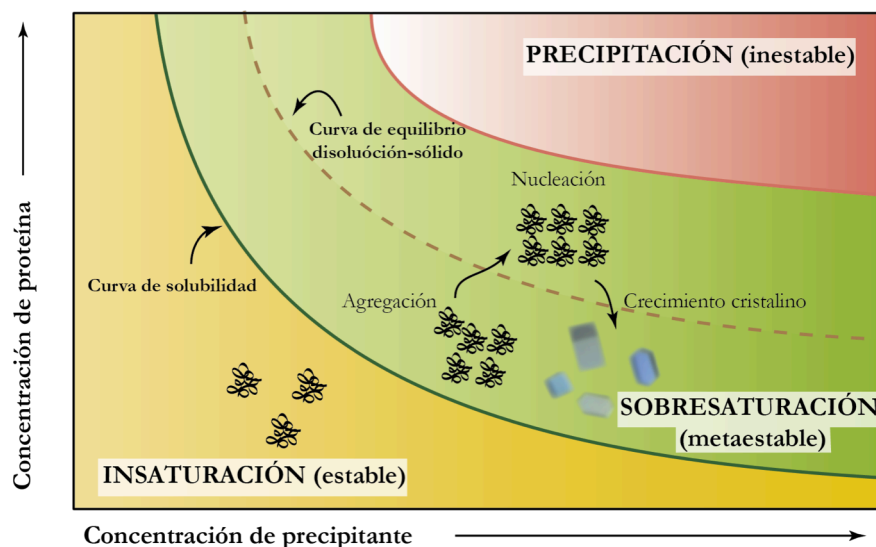
**Figura 3.2. Representación esquemática de unidad asimétrica, celdilla unidad y cristal.** La unidad asimétrica se repite dentro de la celdilla unidad, aplicando los elementos de simetría. A su vez, la celdilla unidad se repite, por traslaciones, formando el cristal tridimensional

El conjunto de operaciones de simetría compatibles con las repeticiones por traslación (retículo cristalino) está limitado a 32 combinaciones diferentes, denominadas *clases cristalinas*, o grupos puntuales cristalinos. Pero a su vez, el número de retículos cristalinos diferentes, compatibles con estas combinaciones no es infinito, sino que está limitado a 14, conocidos con el nombre de *redes de Bravais*.

La combinación de los distintos tipos de retículos posibles (14 *redes de Bravais*) y de los conjuntos de elementos de simetría compatibles (32 *clases cristalinas*) conduce a 230 modos de repetición diferentes mediante los cuales los átomos y las moléculas se empaquetan en el espacio cristalino. De estas 230 modalidades de repetición, que se denominan *grupos espaciales*, las macromoléculas biológicas sólo son compatibles con 65 de ellos. Esta reducción es debida a la quiralidad intrínseca de estas moléculas, por lo que las combinaciones de elementos de simetría de los cristales de estas moléculas no pueden contener centros de inversión. Todos los aspectos de simetría de los cristales, incluyendo los *grupos espaciales*, se encuentran recogidos en las denominadas *Tablas Internacionales de Cristalografía*.

### 3.2.2. FACTORES QUE AFECTAN A LA CRISTALIZACIÓN

La cristalización es el proceso en el que a través de un cambio de fase pasamos de tener una macromolécula en solución acuosa a tenerla como un sólido cristalino. El proceso de cristalización se puede ilustrar mediante un diagrama de fases, en el cual se muestra cómo están relacionadas la concentración de proteína y precipitante. Los cristales de proteína se forman en condiciones de sobresaturación, a medida que nos acercamos a la zona de precipitación se produce una agregación de proteína, se forman pequeños núcleos que sirven como base para el crecimiento de los cristales (Fig. 3.3).



**Figura 3.3. Diagrama de fases.** Se muestran las diferentes zonas en las que la proteína se puede encontrar durante un experimento de cristalización.

Para que tenga lugar la cristalización de una proteína se requiere:

1. Una muestra de proteína pura. La presencia de contaminantes puede evitar la formación de cristales.

2. Que el solvente en el que se encuentra la proteína permita que precipite con una sal o un compuesto orgánico.
3. Que la solución de proteína se acerque a la sobresaturación, donde se produce la formación de pequeños agregados de proteína que sirven como núcleos de cristalización.
4. Una vez que se han formado los núcleos comienza el crecimiento del cristal.

#### ***Solución precipitante.***

**Agente precipitante.** Suelen usarse polímeros de diferente masa molecular, como el polietilenglicol, y sales. También se pueden utilizar compuestos orgánicos volátiles o no, como los alcoholes. Además de la naturaleza del precipitante también se varía su concentración.

**pH.** Afecta a la solubilidad de la proteína, el mínimo de solubilidad se alcanza cerca del punto isoelectrico de la proteína. Se usan rangos de pH en torno a 4,0 y 8,5.

**T<sup>a</sup>.** Se trabaja a una temperatura de 4°C y 18°C.

**Aditivos.** Se emplean sales, alcoholes, azúcares, detergentes y otras moléculas de pequeño tamaño a baja concentración

#### ***Solución de proteína.***

Se suele usar una concentración de proteína en torno a 10 mg/ml. Esta concentración variará en función de la solubilidad y/o estabilidad de la muestra en solución. Así, una proteína muy soluble requerirá trabajar a una mayor concentración para que alcance el estado de sobresaturación, la solubilidad y/o estabilidad pueden verse afectadas por la naturaleza del tampón, el valor del pH y la presencia de sales. Se suelen utilizar tampones a una concentración baja (20- 40 mM) y se evitan aquellos tampones o compuestos que puedan formar cristales y dificultar el proceso de cristalización.



**Tabla 3.1.** Parámetros de cristalización. Factores físicos, químicos y bioquímicos que afectan al experimento de cristalización.

Factores Físicos	Factores Químicos	Factores Bioquímicos
Temperatura	Tipo de precipitante	Pureza de la muestra
Metodología	Concentración del precipitante	Impurezas macromoleculares
Tiempo	pH y tampón	Agregación
Presión	Fuerza iónica	Modificaciones post-traduccionales
Gravedad, convección y sedimentación	Ambiente reductor/ oxidante	Origen de la muestra
Vibraciones y sonido	Concentración de la muestra	Almacenamiento de la muestra
Campos magnéticos	Iones metálicos	Proteólisis
Campos eléctricos	Detergentes	Modificaciones químicas
Propiedades dieléctricas	Impurezas moleculares	Modificaciones en secuencia
Viscosidad	Poli-iones	Simetría de la muestra
Superficie del material de cristalización	Agentes entrelazantes	pI de la muestra
Agentes nucleantes	Metales pesados	Historia de la muestra
Volumen	Origen de los reactivos	Ligandos, cofactores, inhibidores
Material particulado/ amorfo	Pureza de los reactivos	Contaminación microbiana
Manipulación de la muestra	Formulación de los reactivos	Métodos de purificación

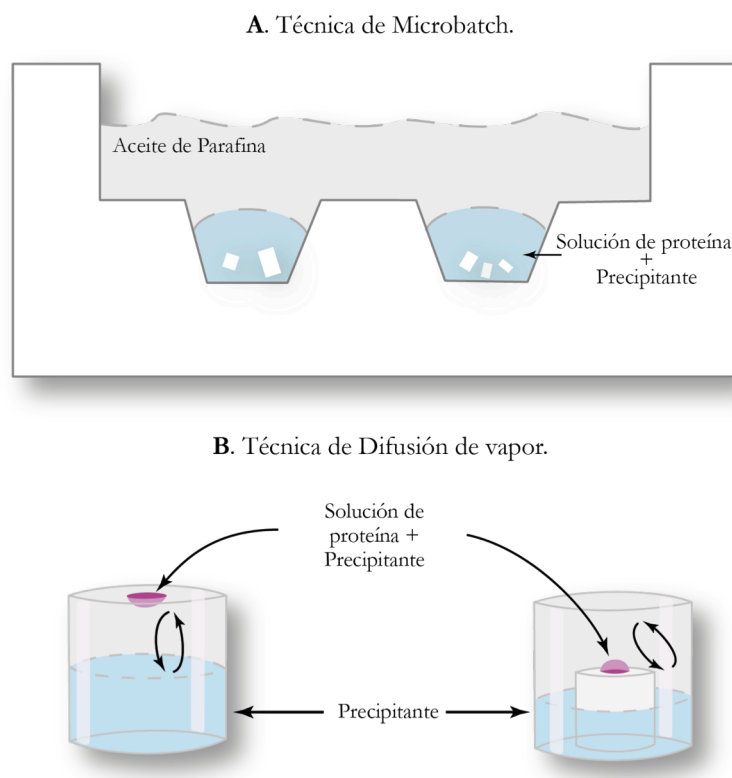
### 3.2.3. BÚSQUEDA DE CONDICIONES DE CRISTALIZACIÓN

En general, encontrar las condiciones adecuadas en las que una proteína cristaliza requiere de muchos intentos, dadas las múltiples variables que intervienen en el proceso. A la hora de cristalizar una proteína nueva, no se conoce, *a priori*, ninguna información que indique cuál será la condición de cristalización más favorable, por lo que es imprescindible llevar a cabo búsquedas sistemáticas de soluciones insolubilizantes más adecuadas. Para ello se suelen usar baterías de soluciones preparadas comercialmente, entre las que destacan Crystal Screen I, Crystal Screen II, Salt RX, e Index de la casa comercial Hampton (Hampton Research, CA, USA), y JCSG+Suite y PACT+Suite de QUIAGEN. En nuestro laboratorio contamos, además, con el robot de cristalización *Gilson-Innovadine* semiautomatizado que permite realizar numerosas pruebas con un coste mínimo de tiempo y cantidad de muestra (tan sólo 250 nL).

Una vez obtenidos algunos cristales iniciales, será necesario poder reproducirlos y optimizarlos, algo que, en muchos casos, puede ser más complicado que la obtención inicial de cristales. Y para ello se suelen utilizar detergentes y aditivos que ayuden a la formación de los mismos y que mejoren su calidad.

### 3.2.4. EXPERIMENTOS DE CRISTALIZACIÓN

Existen varios métodos de cristalización, como son la difusión de vapor, contradifusión, diálisis, microsiembra o macrosiembra. En este trabajo se han empleado principalmente las técnicas de difusión de vapor y microbatch.



**Figura 3.4. Esquema de las diferentes técnicas de cristalización. a) Técnica de *microbatch*. b) Técnica de difusión de vapor en gota colgante (derecha) y gota sentada (izquierda).**

#### *Técnica de Microbatch.*

La técnica de *microbatch* fue desarrollada en 1990 (Chayen, 1998) para llevar a cabo experimentos de cristalización usando cantidades muy pequeñas de proteína. Para evitar la evaporación de la gota, los ensayos se dispensan e incuban bajo una aceite de parafina de baja densidad. Las gotas de cristalización se mantienen bajo el aceite ya que son más densas que el aceite de parafina.

Este método consiste en mezclar la solución de proteína y la solución precipitante a sus concentraciones finales. El agua de las gotas de cristalización se evapora lentamente a través de la capa de aceite, resultando en un incremento de la concentración de proteína y precipitante hasta que se alcanza el punto de nucleación (Fig. 3.4 a).

***Difusión de vapor.***

En este método las gotas se preparan mezclando el mismo volumen de solución de proteína y de precipitante. Es un sistema cerrado que evolucionará por equilibrio de vapor, como la mezcla de proteína y precipitante en la gota está menos concentrada que la solución precipitante en el pocillo, el agua de ésta se evaporará, uniéndose a la solución del pocillo. Como resultado de este proceso, la concentración de la proteína y del precipitante aumentarán lentamente en la gota, y si las condiciones son las adecuadas, se formarán cristales.

El método más común para estos experimentos es el basado en la técnica de la gota colgante. Se colocan unos pocos microlitros (1-2  $\mu\text{L}$ ) de la solución de proteína en el centro de un cubre-objetos de vidrio y se la añaden otros tantos microlitros de la solución precipitante. El cubre-objetos que contiene la gota de mezcla se le da la vuelta y con él se tapa el pocillo mencionado (Fig. 3.4 b).

En el método de gota sentada la mezcla de solución de proteína y precipitante se sitúa sobre una elevación plana o cóncava que se encuentra por encima del nivel de la solución de cristalización, finalmente se sella con una lámina flexible de plástico adhesivo y transparente (Fig. 3.4 b).

**3.2.5. CRISTALIZACIÓN DE COMPLEJOS DE PROTEÍNA-PROTEÍNA**

Para la cristalización de los complejos de proteína-ligando se pueden llevar a cabo dos procedimientos diferentes: la inmersión (*soaking*) o co-cristalización (Hassell *et al.*, 2007).

*Co-cristalización.* La proteína se incuba con el ligando previamente o se mezclan directamente en la gota de cristalización, así la proteína cristalizará en presencia del ligando. En algunos casos la incubación previa permite que zonas móviles se estabilicen y la proteína se empaquete correctamente.

*Inmersión (soaking).* Los cristales de proteína se sumergen en una solución con ligando. El ligando difundirá por los canales de solvente presentes en el cristal hasta que se forme el complejo proteína-ligando. Para ello es necesario que el sitio activo sea accesible y que el empaquetamiento de los cristales no sea muy denso, dejando acceso al ligando. Por lo tanto, este método sólo se utiliza en el caso de ligandos de pequeño tamaño, que sean capaces de pasar entre los canales de solvente.

**3.2.6. PREPARACIÓN DE LAS MUESTRAS**

La preparación de las muestras requiere de tres pasos: crio-protección, enfriamiento rápido (en inglés *shock-cooling*) y el trasvase del cristal a la cabeza goniométrica del aparato de rayos X.

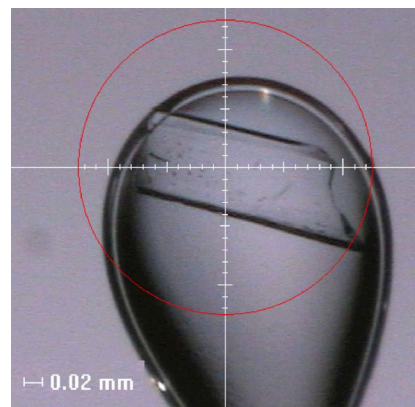
La crioprotección de los cristales requiere de la introducción de agentes crioprotectores tanto dentro como fuera del cristal (el cual contiene entre 40- 90% de solvente) y de la eliminación del agua del exterior del mismo.

Al ser expuestos a los rayos X los cristales sufren un importante daño, producto de la radiación, que afecta a la estructura interna y que causa la pérdida del poder de difracción. Para minimizar estos efectos indeseados y alargar la vida útil del cristal, maximizando el tiempo total de exposición, los cristales son vitrificados en nitrógeno líquido. Durante el proceso de vitrificación es muy importante evitar la formación de micro-cristales de hielo, que interfieren en la difracción propia del cristal y producen la disrupción del mismo. Por este motivo, los cristales se sumergen previamente en una solución que contiene un agente crio-protector junto con el líquido madre. El agente crioprotector difundirá sobre la superficie del cristal e incluso en algunos casos penetrará al mismo, protegiéndolo y permitiendo el adecuado proceso de vitrificación.

Los agentes crioprotectores más usados son: PEG, MPD, glicerol, paratona, criosales, etc. Los crioprotectores más comunes son efectivos en una concentración entre 15 y 25% (v/v). Las pruebas de crioprotección se realizan siempre primero a las concentraciones más bajas posibles, incrementando esta gradualmente a los efectos de utilizar la mínima concentración de agente crioprotector.

Para poder ser transportados y manipulados fácilmente los cristales son pescados de la gota de cristalización con la ayuda de un pequeño lazo de nylon unido a una base metálica magnética (Fig. 3.5).

Después de montar los cristales se sumergen durante unos segundos en una gota que contiene tanto la solución del reservorio como el agente crioprotector a una concentración adecuada. Por último, el lazo se sumerge rápidamente en nitrógeno líquido.

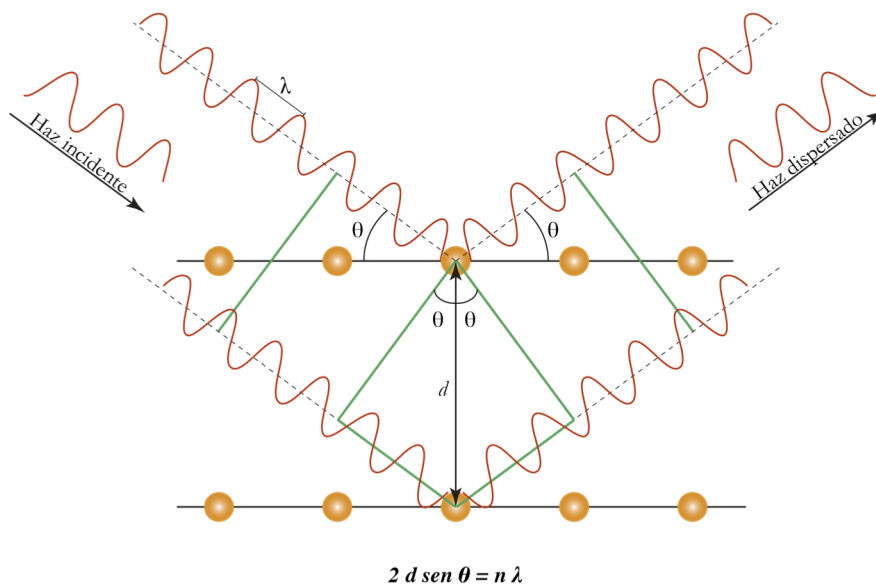


**Figura 3.5.** Fotografía de un cristal montado en un lazo de nylon.

### 3.3. DIFRACCIÓN DE RAYOS X Y RESOLUCIÓN ESTRUCTURAL

La Difracción de Rayos X es el fenómeno físico fundamental a través del cual se manifiesta la interacción de los rayos X con los cristales. Cuando un haz de rayos X incide sobre un cristal, éste interacciona con los electrones de los átomos que componen el cristal, haciéndolos vibrar acopladamente con las variaciones periódicas de su campo eléctrico. Los electrones oscilantes se convierten así en focos de nueva radiación X que se emite en forma esférica, fenómeno denominado *dispersión*, normalmente elástica. Los rayos X dispersados por los electrones interfieren entre sí, pudiendo dar lugar a cancelaciones. Sin embargo, como los átomos en un cristal están ordenados de manera regular y periódica, también ocurre que, en determinadas direcciones, se combinan y refuerzan para dar lugar a lo que se denomina dispersión cooperativa o *difracción*. En el caso de que se de una interferencia constructiva se cumple la ley de Bragg, la cual describe el fenómeno de difracción (Fig. 3.6). El uso de la radiación X se

debe a que el fenómeno de difracción requiere que la longitud de onda de la radiación electromagnética sea del mismo orden de las distancias entre los focos dispersores, los átomos del cristal.



**Figura 3.6. Ley de Bragg.** Cuando los frentes de onda emergentes tras la reflexión están en fase, se observará intensidad, es decir, se estará cumpliendo la ley de Bragg.

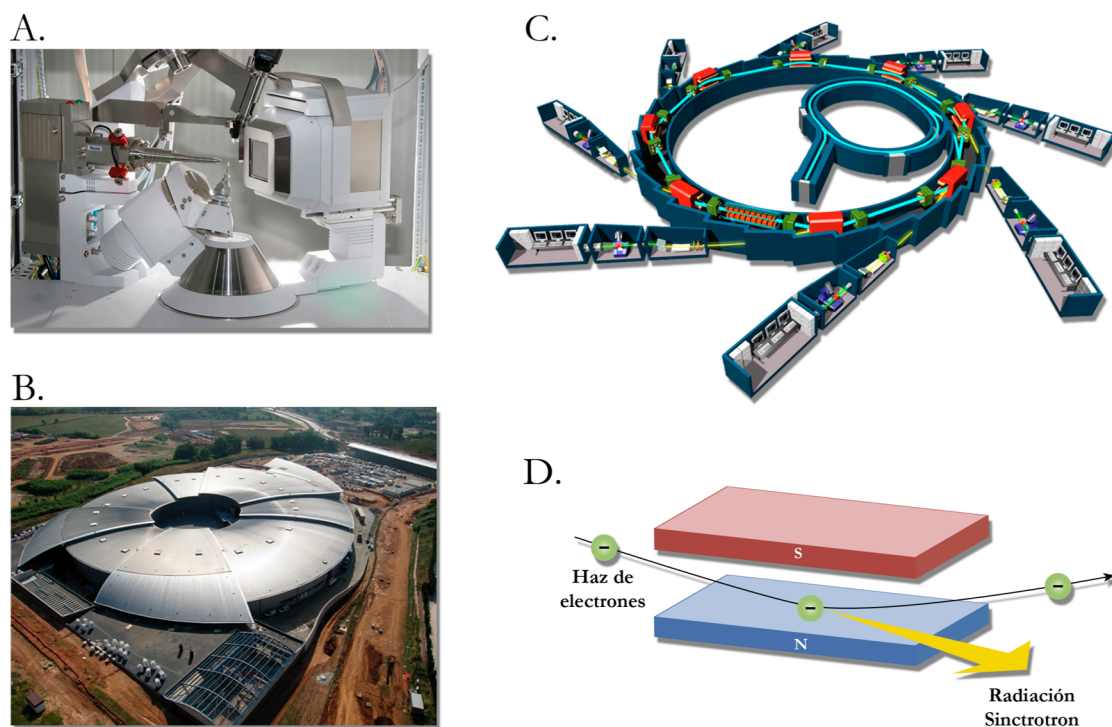
El modo en que la cristalografía de rayos X llega al conocimiento de la estructura cristalina de cualquier material puede comprenderse mejor si se compara con el símil que representa la observación de un objeto a través de un microscopio óptico. En éste, la radiación dispersada por el objeto en estudio se recombina de nuevo, a través de un sistema de lentes, para dar lugar a la imagen aumentada del objeto dispersor. Sin embargo, no existe ninguna lente que sea capaz de focalizar los rayos X dispersados por los átomos. Para ello, las técnicas cristalográficas hacen uso de cálculos matemáticos “síntesis de Fourier”, con los que se consigue focalizar los rayos X.

### 3.3.1. FUENTES DE RADIACIÓN

Los rayos X que más interesan en el campo de la Cristalografía son aquellos que disponen de una longitud de onda alrededor de 1 Å, pues esa longitud de onda es muy próxima a las distancias entre los átomos. Este tipo de radiación X se produce en los laboratorios de Cristalografía mediante generadores de ánodo rotatorio o en las llamadas grandes instalaciones de sincrotrón. En este trabajo se ha utilizado principalmente las fuentes de radiación sincrotrón.

Los equipos que se utilizan en los laboratorios de Cristalografía para producir estos rayos X son relativamente sencillos. Disponen de un generador de alta tensión (unos 50.000 voltios), que se suministra al llamado tubo de rayos X, que es realmente donde se produce la radiación. Esos 50 kV se suministran como diferencia de potencial (alto voltaje) entre un filamento incandescente (por el que se hace pasar una corriente de bajo voltaje, unos 5 A a unos 12 V) y un metal puro (normalmente cobre o

molibdeno), estableciéndose entre ambos una corriente de unos 30 mA de electrones libres. Desde el filamento incandescente (cargado negativamente) saltan electrones hacia el ánodo (cargado positivamente) provocando, en los átomos de este último, una reorganización electrónica en sus niveles de energía.



**Figura 3.7. Radiación sincrotrón.** a) Detalle de un generador de rayos X y su detector. b) Imagen aérea de las instalaciones del sincrotrón ALBA en Barcelona. c) Esquema general de un sincrotrón. d) Sistema de imanes para la aceleración y redireccionamiento de las partículas internas de un sincrotrón.

Este es un proceso en el que se genera mucho calor, por lo que los tubos de rayos X deben estar muy refrigerados. Una alternativa a los tubos convencionales son los llamados *generadores de ánodo rotatorio*, en los cuales el ánodo, en forma de cilindro, se mantiene con un giro continuo, consiguiendo con ello que la incidencia de los electrones se reparta por la superficie del cilindro y así se puedan obtener potencias mayores de rayos X (Fig. 3.7 a).

**Radiación sincrotrón.** Una instalación sincrotrónica contiene un anillo por el que se hacen circular electrones a altísima velocidad en el interior de canales rectilíneos que de vez en cuando se quiebran para adaptarse a la curvatura del anillo (Fig. 3.7 b y 3.7 c). A estos electrones se les hace cambiar de dirección para pasar de un canal a otro usando campos magnéticos de gran energía (Fig. 3.7 d), y es en ese momento, cuando los electrones emiten una radiación.

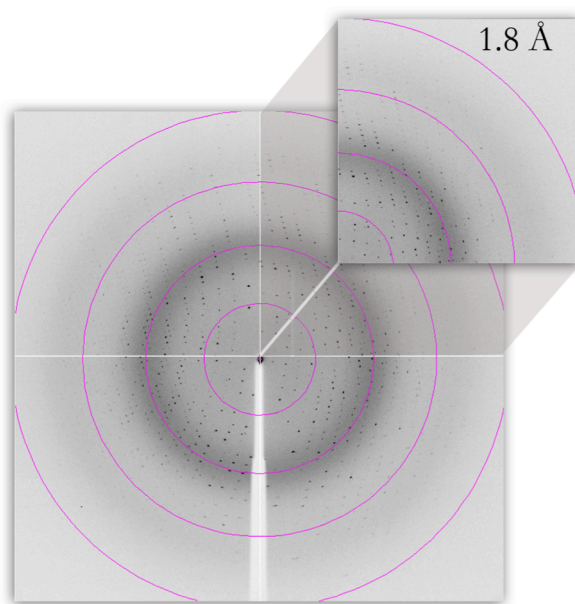


Los rayos X obtenidos en las instalaciones de sincrotrón tienen dos grandes ventajas para la Cristalografía de rayos X: la longitud de onda se puede modular a voluntad, muy útil a la hora de resolver el problema de las fases necesarias para la resolución estructural; y su brillo es, como mínimo, un billón de veces ( $10^{12}$ ) superior a la de los rayos X convencionales.

Las medidas de difracción tomadas a lo largo de este trabajo de tesis se han llevado a cabo en los sincrotrones ESRF (Grenoble, Francia), SLS (Villigen, Suiza) y ALBA (Barcelona, España).

### 3.3.2. PROCESADO DE DATOS DE DIFRACCIÓN

Una vez que se ha montado el cristal, éste es irradiado con rayos X mientras se le hace girar en intervalos adecuados. La medida realizada en los sucesivos intervalos nos proporciona un conjunto de imágenes de difracción en las que se recoge la totalidad de los haces difractados por el cristal, los cuales constituyen el patrón de intensidades (Fig. 3.8). A partir del espectro de difracción podremos obtener las dimensiones de la celdilla unidad, que se calculan a partir de la posición relativa de los distintos haces difractados, y además el grupo espacial, que se obtiene del análisis de las intensidades que están ausentes.



**Figura 3.8. Imagen de un patrón de difracción de proteína.** Imagen de difracción recogida en una recolección de datos de un cristal de proteína. Arriba a la derecha se muestra un detalle del patrón de difracción con el valor de la máxima resolución.

El procesamiento de datos de difracción consta de tres pasos: el indexado, la integración y el escalado; existen programas específicos que permiten realizar los cálculos requeridos en dichos pasos.

**Indexado.** Es el primer paso para la extracción de información contenida en las imágenes de difracción e implica asignar los correspondientes índices de Miller a cada una de las reflexiones observadas en la misma.

Los índices de Miller son tres números enteros hkl que identifican familias de planos paralelos que pueden definirse en el cristal y que de acuerdo a la construcción geométrica de W. H. y W. L. Bragg están asociados con la difracción de la red cristalina.

Para el indexado se utilizaron los programas MOSFLM/ XDS que calculan, en primer lugar, la matriz de orientación del cristal, y en función de ella establecen los posibles grupos espaciales, la celda unidad y asignan los índices de Miller.

**Integración.** Una vez que es posible asociar cada una de las reflexiones con sus correspondientes índices de Miller, se procede a la integración. La integración permite calcular la intensidad de cada una de las reflexiones observadas.

**Escalado.** Debido a que durante la recogida de datos la intensidad del haz de rayos X varía, los datos deben ser recogidos y puestos en la misma escala. Se lleva a cabo la corrección de las intensidades teniendo en cuenta el ruido de fondo de las imágenes, la mosaicidad del cristal, el decaimiento de la difracción del cristal y las fluctuaciones en la intensidad de la fuente de rayos X. Este procedimiento se conoce con el nombre de escalado de datos. Para el mismo se empleó el programa SCALA del paquete informático CCP4 (*Collaborative Computational Project*, 1994) (Bale, 1994). Algunos de los criterios de la calidad de los datos de difracción son:

$R_{sym}$ . Indica las diferencias en la intensidad de las reflexiones relacionadas por simetría. Los valores globales para un buen conjunto de datos deben ser menores del 10%

$I/\sigma(I)$ . Representa la intensidad de la señal respecto del ruido. Son aceptables valores por encima de 2.

**Multiplidad o redundancia.** Es el número de veces que, en promedio, se ha medido cada reflexión. Cuanto mayor es su valor, mejor será la estimación de los errores asociados a la medida de la reflexión.

**Complejidad.** Indica el porcentaje del espacio recíproco (espectro) que ha sido medido. La complejidad global necesaria comprende un 90- 100%, así como del 80% en la última capa de resolución.

**Cálculo del módulo de factor de estructura.** El factor de estructura es la onda resultante del haz difractado por todos los átomos de la celdilla en una determinada dirección. Como cualquier onda, el factor de estructura viene caracterizado por su módulo,  $|F_{hkl}|$  y su fase,  $\alpha_{hkl}$ . El módulo representa la



amplitud del haz difractado y puede calcularse directamente a partir de la intensidad medida para dicha reflexión, pero la información de la fase, en cambio, se pierde durante el experimento de difracción. La transformación de I a F se hizo mediante el programa TRUNCATE del paquete informático CCP4 (*Collaborative Computational Project*, 1994) (Baley, 1994).

### 3.3.3. MÉTODOS DE RESOLUCIÓN ESTRUCTURAL

El proceso de resolución estructural implica la determinación, a partir de las intensidades de difracción experimentalmente obtenidas, de la función de densidad electrónica  $\rho(x,y,z)$  correspondiente. Esta función nos permite conocer las posiciones atómicas de cada uno de los átomos que constituyen la molécula proteica. La función de densidad electrónica está dada por la siguiente relación:

$$\rho(xyz) = \frac{1}{V} \sum_{hkl}^{+\infty}_{-\infty} |F(hkl)| \cdot e^{-2\pi i(hk+lz-\phi(hkl))}$$

**Ecuación 1. Función de densidad electrónica.** V es el volumen de la celda unidad, x, y, z las coordenadas de cualquier punto del espacio, hkl los índices de Miller,  $|F(hkl)|$  el módulo del factor de estructura y  $\Phi(hkl)$  su correspondiente fase.

Los factores de estructura representan la contribución a la difracción, para cada familia de planos (hkl), de los átomos contenidos en la celda unidad. Estos factores de estructura son la suma de una serie de términos que dependen básicamente de tres parámetros: el tipo de átomo, su poder de dispersión de los rayos X (descrito mediante el factor de dispersión atómica  $f_j$ ) y la posición del átomo en la celda unidad.

$$F(hkl) = \sum_j f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

**Ecuación 2.** Factor de estructura de cada haz difractado. Recoge la dispersión f de todos los electrones de todos los átomos j que están contenidos en la celdilla elemental del cristal, para la reflexión de índices h, k, l.

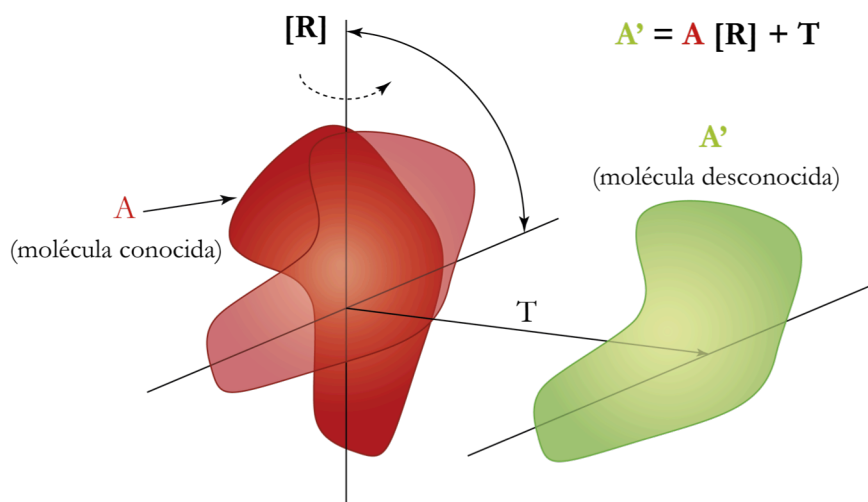
Las intensidades de los haces difractados, obtenidas experimentalmente cuando recogemos las imágenes de difracción, son proporcionales al cuadrado de los módulos de los factores de estructura. Sin embargo, la información de las fases  $\Phi(hkl)$  de los factores de estructura no se puede medir directamente. En cristalografía, este problema se conoce como el “problema de las fases”, y resolverlo, es el punto clave para lograr la determinación estructural.

El “problema de las fases” puede resolverse mediante diversos métodos, los más comunes son: introduciendo átomos altamente dispersores, o método de **Reemplazo Isomorfo Múltiple (MIR)**, del inglés *Multiple Isomorphous Replacement*), introduciendo átomos dispersores anómalos o método de

**Dispersión Anómala Simple o Múltiple (SAD, del inglés *Single-wavelength Anomalous Diffraction* o MAD, del inglés *Multi-wavelength Anomalous Diffraction*), y mediante Reemplazo Molecular (MR, del inglés *Molecular Replacement*), haciendo uso del modelo estructural de una proteína homóloga, previamente determinada. Con cualquiera de estos tres métodos podemos calcular unas fases iniciales, que junto con las amplitudes experimentales, nos permitirán calcular una función de densidad electrónica aproximada, y sobre la cual podremos construir nuestro modelo estructural. En este trabajo se ha utilizado el método MR como método de resolución estructural.**

### *El método MR*

Cuando se dispone de un modelo estructural de una proteína con secuencia de aminoácidos homóloga, el problema de las fases se puede resolver mediante la técnica de del reemplazo molecular (MR). La estructura de esta proteína homóloga se considera como si fuera la proteína que va a determinarse y sirve como un primer modelo que posteriormente será refinado. Este procedimiento está basado en la observación de que proteínas homólogas en su secuencia peptídica muestran un plegamiento muy similar. El problema consiste en transferir la estructura molecular de la proteína conocida, desde su propio empaquetamiento cristalino hasta el cristal de la proteína con estructura desconocida. El posicionamiento de la molécula conocida en la celdilla unidad de la proteína desconocida requiere determinar su correcta orientación y su posición precisa. Ambas operaciones, rotación y traslación, se calculan mediante las funciones denominadas de rotación y traslación (Fig. 3.9).



**Figura 3.9. Esquema del proceso de reemplazo molecular (MR).** La molécula de estructura conocida (A) se gira mediante la operación de rotación [R] y se traslada mediante T, para llevarla a la posición que ocupa la molécula (A').

La calidad de las soluciones de estas funciones se expresa mediante los coeficientes de correlación entre las funciones de Patterson experimental y la calculada con la proteína conocida. Un coeficiente de

correlación alto entre dichas funciones equivale a un buen acuerdo entre el espectro de difracción experimental y el calculado con la proteína conocida. Una vez orientada y trasladada convenientemente la molécula conocida, se calcula un mapa de densidad electrónica usando los factores de estructura experimentales.

### 3.3.4. REFINAMIENTO DE LOS MODELOS.

Una vez solucionado el problema de las fases, utilizando cualquiera de los métodos comentados en el apartado anterior, es posible calcular el primer mapa de densidad electrónica, que no es más que la representación gráfica de la función de densidad electrónica.

En ausencia de un modelo estructural homólogo la construcción del modelo comienza por el trazado de la cadena polipeptídica dentro del mapa de densidad electrónica. Esta labor puede hacerse manualmente o bien, de forma automática, utilizando programas gráficos que permiten identificar los máximos de densidad electrónica en los que se situarán los átomos de nuestra molécula. Cuando se utiliza el Reemplazo Molecular como método de faseado, ya tenemos gran parte de la cadena polipeptídica trazada en el mapa de densidad, la del modelo estructural utilizado.

Una vez trazada la cadena polipeptídica ya tenemos un modelo inicial. Pero éste aún es insuficiente para definir la estructura de nuestra proteína, por lo que ha de mejorarse mediante sucesivas etapas de refinamiento y modelado manual.

#### *Refinamiento matemático.*

Los parámetros que se refinan en una estructura cristalográfica son:

- Coordenadas atómicas (x,y,z): determinan la posición de cada átomo dentro de la celdilla unidad.
- Factor térmico B: representa la vibración térmica o movilidad de cada átomo.
- Ocupación: hace referencia a la fracción de moléculas en las que un átomo ocupa una determinada posición.

El tamaño de las moléculas de proteína hace que el número de parámetros a refinar sea del orden del número de observaciones, por lo que es necesario introducir una serie de restricciones geométricas (distancias y ángulos de enlace, torsiones, etc.), que en el caso de los aminoácidos y otras moléculas son de gran precisión.

El método de refinamiento a utilizar depende del estado del modelo. En las primeras etapas del refinamiento se aplican métodos de refinamiento de Cuerpo Rígido (*Rigid body*) y Calentamiento Simulado (*Simulated Annealing*), mientras que en etapas posteriores se utilizan los Métodos Probabilísticos de Máxima Verosimilitud (*Maximum Likelihood*).

Durante el proceso de refinamiento debemos asegurarnos de que los cambios introducidos en el modelo converjan hacia la estructura real de nuestra molécula. Para seguir esta convergencia calculamos el factor de desacuerdo R, que nos indica la bondad del modelo mediante el cálculo de las diferencias entre el espectro observado ( $|F_o|$ ) y el calculado a partir del modelo experimental ( $|F_c|$ ).

El factor R debe ir disminuyendo a lo largo del refinamiento, pero un valor bajo no justifica siempre que el modelo sea correcto. Para asegurarnos que los cambios que realizamos sobre el modelo son representativos de la estructura real, utilizamos un Método Estadístico de Validación Cruzada basado en el cálculo de un segundo factor de convergencia, el  $R_{\text{free}}$ . Este método implica la separación del total de las reflexiones en dos conjuntos, un 90-95% de las reflexiones que serán utilizadas en el refinamiento (*working test*) y un 5-10% de las reflexiones que se emplearán solo para el control (*test set*). Por lo tanto, el factor  $R_{\text{free}}$  indica el desacuerdo que existe entre el modelo refinado y las reflexiones que no utilizamos en el refinamiento, por lo que solo los cambios que disminuyan el  $R_{\text{free}}$  serán correctos.

$$R = \frac{\sum_{hkl} |F_{\text{obs}}| - k |F_{\text{cal}}|}{\sum_{hkl} |F_{\text{obs}}|} \times 100$$

**Ecuación 3.** Fórmula del factor de desacuerdo R.

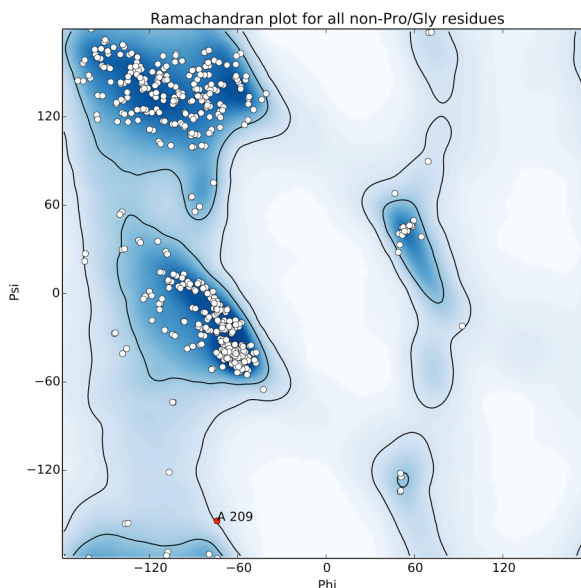
Para el refinamiento Matemático se emplearon los programas REFMAC (Murshudov *et al.*, 1997) y PHENIX (Adams *et al.*, 2010).

### 3.3.5. MODELADO MANUAL Y VALIDACIÓN DEL MODELO FINAL.

Algunos de los sesgos del modelo no pueden solventarse mediante refinamiento matemático y han de corregirse manualmente mediante métodos visuales. Estos sesgos se deben a:

- Errores en la colocación de la cadena principal o lateral de los aminoácidos.
- Errores de secuencia causados por diferencias entre el modelo de Reemplazo Molecular y nuestra molécula.
- Presencia de ligandos ausentes en el modelo inicial.

La corrección de estos errores se realiza sobre los mapas de densidad electrónica calculados mediante una síntesis de Fourier, con los factores de estructura observados ( $F_o$ ) y los factores de estructura calculados para el modelo ( $F_c$ ), y en donde los coeficientes utilizados en el cálculo de los mapas son  $2F_o - F_c$  y  $F_o - F_c$ .



**Figura 3.10. Validación estructural.** Diagrama de Ramachandran donde se muestran las zonas de dispersión aceptable para los valores de los ángulos de torsión de los enlaces peptídicos en un modelo estructural de una macromolécula, dependiendo de las diferentes zonas estructurales (alfa-hélices, lámina-beta, etc.).

El refinamiento se da por finalizado cuando el modelo se ajusta al mapa de densidad electrónica y los factores  $R$  y  $R_{\text{free}}$ , se encuentran en unos valores cercanos o inferiores al 20% ( $R$ ) y menores de 30% ( $R_{\text{free}}$ ). En este punto se debe comprobar que los parámetros tales como las distancias y los ángulos de enlace, la conformación de los rotámeros y los ángulos de torsión sean electroquímicamente aceptables. El diagrama de Ramachandran representa los ángulos de torsión  $\Psi$  y  $\phi$  de los residuos de la cadena polipeptídica (Fig. 3.10), lo que nos permite conocer si los enlaces entre aminoácidos se encuentran dentro del intervalo permitido. Estos parámetros son analizados con los programas PROCHECK (Laskowski *et al.*, 1996) y COOT (Emsley & Cowtan, 2004).

### 3.4. OTROS PROGRAMAS UTILIZADOS

**Alineamiento de secuencias.** Se emplearon los programas BLAST (Altschul & Lipman, 1990) y CLUSTALW (Thompson *et al.*, 1994).

**Análisis de la estructura secundaria.** El análisis de la estructura secundaria de los modelos estructurales obtenidos se llevó a cabo a través del programa PROCHECK (Laskowski *et al.*, 1996) que revisa la estereoquímica de la molécula.

**Representación gráfica y elaboración de figuras.** Para la elaboración de las figuras se utilizó el programa PYMOL (DeLano, 2002).

**Servidor Dali.** A través de este servidor se pueden comparar estructuras de proteínas previamente resueltas y depositadas en el PDB (Holm & Rosenstrom, 2010).



## 4. TRABAJOS DE INVESTIGACIÓN





# Trabajo de Investigación I:

## Crystallization and preliminary X-ray diffraction analysis of phosphoglycerate kinase from *Streptococcus pneumoniae*

Noelia Bernardo-García, Sergio G. Bartual, Marcus Fulde, Simone Bergmann and Juan A. Hermoso

*Acta Crystallographica Section F* (2011). **F67**, 1285-1289



#### **4.1. INTRODUCCIÓN A LA CRISTALIZACIÓN DE LA FOSFOGLICERATO QUINASA Y APORTACIÓN DEL AUTOR**

La resolución estructural de la molécula para el posterior análisis de la estructura requiere de un paso previo fundamental, la obtención de cristales de proteína.

La fosfoglicerato quinasa (PGK) es una enzima ampliamente distribuida que participa en la ruta de la glicólisis. Algunas enzimas de esta ruta, como la enolasa o la GADPH, se han identificado como proteínas de superficie en neumococo involucradas en virulencia y en la unión a proteínas del hospedador. Esta unión permitiría a las células del neumococo diseminarse con facilidad a través del epitelio y el endotelio.

La PGK (42 kDa) cataliza la segunda etapa de la glicólisis, la conversión reversible de 1,3-difosfoglicerato (BPG) a 3-fosfoglicerato (3PG) con la generación de una molécula de ATP.

A través de este trabajo nos planteamos los siguientes objetivos:

- ✓ Establecer el protocolo de expresión y purificación de PGK.
- ✓ Conocer las condiciones óptimas de cristalización de PGK.
- ✓ Obtener las estadísticas de datos de difracción preliminares empleados para resolver la estructura cristalográfica de PGK.

La autora de esta Tesis llevo a cabo la cristalización de la proteína, así como la toma de los datos en el sincrotrón ESRF (Grenoble, Francia) y del procesado de los mismos.



**Crystallization and preliminary X-ray diffraction analysis of phosphoglycerate kinase from *Streptococcus pneumoniae***

Noelia Bernardo-García, Sergio G. Bartual, Marcus Fulde, Simone Bergmann y Juan A. Hermoso

Acta Crystallographica Section F (2011) F **67**, 1285- 1289



**Noelia Bernardo-García,<sup>a</sup>  
Sergio G. Bartual,<sup>a</sup> Marcus  
Fulde,<sup>b</sup> Simone Bergmann<sup>b,c</sup> and  
Juan A. Hermoso<sup>a\*</sup>**<sup>a</sup>Department of Crystallography and  
Structural Biology, Instituto de Química-Física  
'Rocasolano', CSIC, Serrano 119,  
28006 Madrid, Spain, <sup>b</sup>Department of  
Medical Microbiology, Helmholtz Centre of  
Infection Research, Inhoffenstrasse 7,  
38124 Braunschweig, Germany, and  
<sup>c</sup>Department of Infection Biology, Institute for  
Microbiology, Technische Universität  
Braunschweig, Spielmannstrasse 7,  
38106 Braunschweig, Germany

Correspondence e-mail: xjuan@iqfr.csic.es

Received 26 May 2011  
Accepted 1 August 2011

## Crystallization and preliminary X-ray diffraction analysis of phosphoglycerate kinase from *Streptococcus pneumoniae*

Phosphoglycerate kinase (PGK) is a widespread two-domain enzyme that plays a critical role in the glycolytic pathway. Several glycolytic enzymes from streptococci have been identified as surface-exposed proteins that are involved in streptococcal virulence by their ability to bind host proteins. This binding allows pneumococcal cells to disseminate through the epithelial and endothelial layers. Crystallization of PGK from *Streptococcus pneumoniae* yielded orthorhombic crystals (space group *I*222, unit-cell parameters  $a = 62.73$ ,  $b = 75.38$ ,  $c = 83.63$  Å). However, the unit cell of these crystals was not compatible with the presence of full-length PGK. Various analytical methods showed that only the N-terminal domain of PGK was present in the *I*222 crystals. The ternary complex of PGK with adenylyl imidodiphosphate (AMP-PNP) and 3-phospho-D-glycerate (3PGA) produced monoclinic crystals (space group *P*2<sub>1</sub>, unit-cell parameters  $a = 40.35$ ,  $b = 78.23$ ,  $c = 59.03$  Å,  $\beta = 96.34^\circ$ ). Molecular replacement showed that this new crystal form contained full-length PGK, thereby indicating the relevance of including substrates in order to avoid proteolysis during the crystallization process.

### 1. Introduction

*Streptococcus pneumoniae* is a major cause of life-threatening infections such as pneumonia, meningitis and septicaemia worldwide. Successful infection depends on pathogen–host interaction and it is thought that surface-exposed proteins are critical to this relationship by providing the first contact of the pathogen with its intended host (Bergmann & Hammerschmidt, 2006; Hammerschmidt, 2006). Recently, it has been shown that streptococcal phosphoglycerate kinase (PGK), a member of the glycolytic proteins, is a surface-exposed protein. Proteomic analyses of *S. agalactiae* have revealed the presence of PGK, enolase and glyceraldehyde-3-phosphodehydrogenase (GAPDH) on the surface of these group B streptococci (Hughes *et al.*, 2002). However, the mechanisms of cell-wall binding and secretion for these glycolytic proteins remain unknown (Bergmann *et al.*, 2003). Glycolytic enzymes of streptococci have been identified as multifunctional virulence factors, mediating interaction with host components involved in haemostasis and in bacterial binding to extracellular matrix proteins (Bergmann & Hammerschmidt, 2006; Pancholi *et al.*, 2003). One example is GAPDH of group A streptococci, which is directly involved in streptococcal adherence and colonization (Jin *et al.*, 2005), in circumventing the host immune defence (Terao *et al.*, 2006) and, together with pneumococcal enolase, in mediating the degradation of extracellular matrix proteins (Bergmann *et al.*, 2005; Hammerschmidt, 2006). *S. pneumoniae* PGK (41.9 kDa) is a major component of the glycolytic pathway, facilitating the transfer of the anhydride phosphate of 1,3-bisphospho-D-glycerate (1,3-BPG) to adenosine 5'-diphosphate magnesium salt (MgADP), yielding 3-phospho-D-glycerate (3PGA) and adenosine 5'-triphosphate magnesium salt (MgATP) (Bernstein *et al.*, 1997). PGK is monomeric and folds into two globular domains (Pecorari *et al.*, 1996). The N-terminal domain binds 3PGA and the C-terminal



© 2011 International Union of Crystallography  
All rights reserved



## crystallization communications

domain binds MgADP; a linker is retained between the two domains that is susceptible to attack by proteases.

Here, we describe the preliminary crystallographic results obtained for the N-terminal domain of PGK (N-PGK) and for full-length PGK as a ternary complex with 3PGA and AMP-PNP.

### 2. Experimental procedures

#### 2.1. Expression and purification of PGK

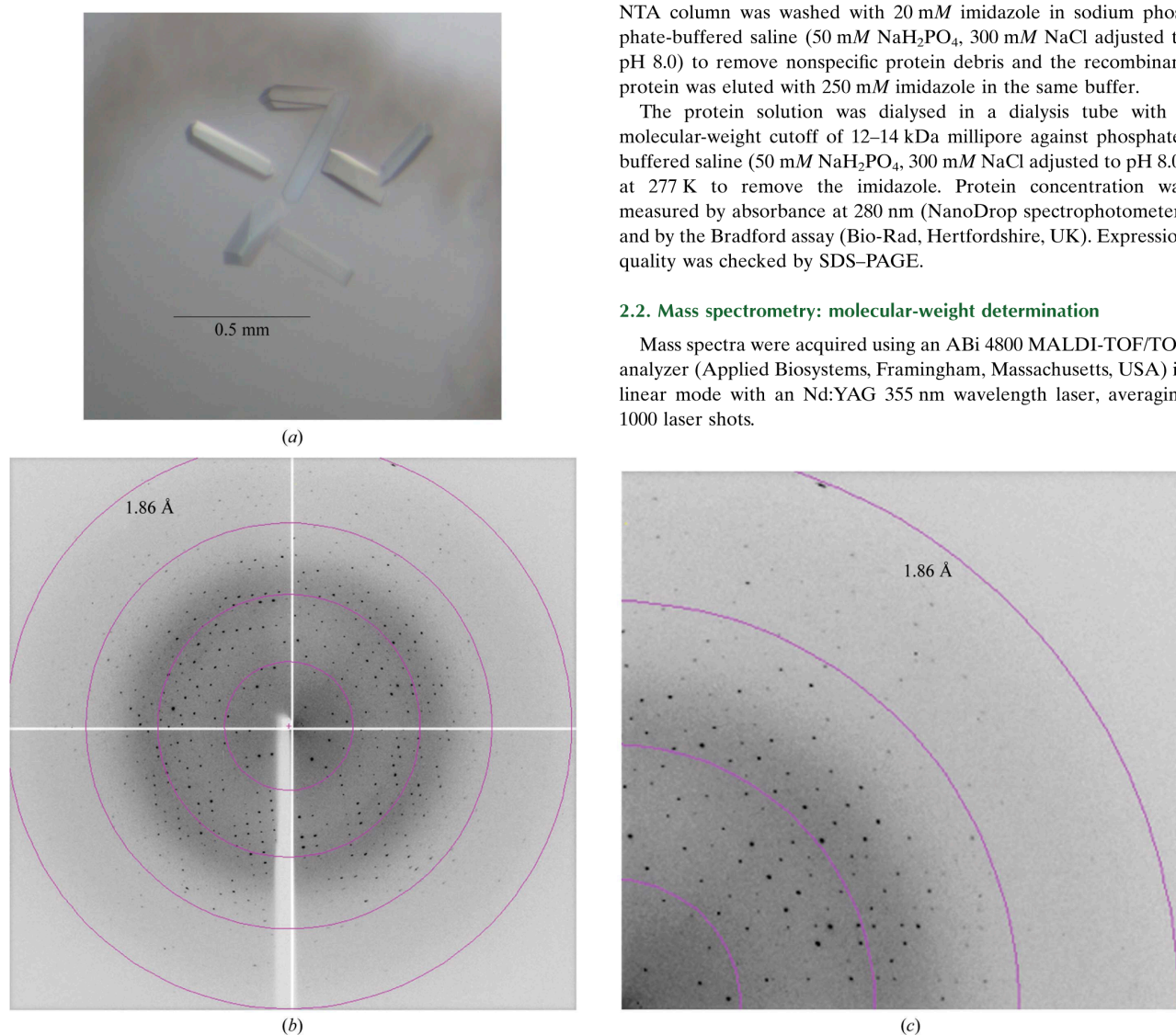
A DNA fragment comprising the full-length protein (YP\_815953.1) was amplified by PCR using the oligonucleotides left primer (5'-GGGGGATCCTTGGCAAACTTACTGTAAAGAG-3') and right primer (5'-CCCGTCTGACTTATTTTCTGTCAAGGCTGCAAG-3'). The sequences in bold indicate restriction-site linkers (*Bam*HI and *Sal*I, respectively) in the pQE30 plasmid (Qiagen,

Düsseldorf, Germany), yielding pQE30PGKw. A six-histidine tag was included at the N-terminus of the construct. A 5 ml aliquot of *Escherichia coli* strain M15 (pREP) freshly transformed with pQE30PGKw was used as a starter inoculum in 250 ml LB medium supplemented with 100 µg ml<sup>-1</sup> ampicillin and 50 µg ml<sup>-1</sup> kanamycin. It was then incubated with vigorous shaking overnight at 309 K. After supplementation with 250 ml LB medium without antibiotics, protein expression was induced by the addition of 1 mM IPTG and growth was continued for 4 h at 301 K. Bacterial cultures were sedimented by centrifugation at 2980g for 20 min. The supernatant was discarded and the pellet was dissolved in 10 ml phosphate-buffered saline (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl adjusted to pH 8.0) supplemented with 1 mM PMSF to reduce nonspecific protease cleavage. Mechanical lysis was performed using a French pressure cell (SLM-Aminco). Lysates were centrifuged for 30 min at 17 211g and the supernatant was loaded onto a Ni-NTA column (Qiagen) for purification of the His-tagged protein at room temperature. The Ni-NTA column was washed with 20 mM imidazole in sodium phosphate-buffered saline (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl adjusted to pH 8.0) to remove nonspecific protein debris and the recombinant protein was eluted with 250 mM imidazole in the same buffer.

The protein solution was dialysed in a dialysis tube with a molecular-weight cutoff of 12–14 kDa millipore against phosphate-buffered saline (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl adjusted to pH 8.0) at 277 K to remove the imidazole. Protein concentration was measured by absorbance at 280 nm (NanoDrop spectrophotometer) and by the Bradford assay (Bio-Rad, Hertfordshire, UK). Expression quality was checked by SDS-PAGE.

#### 2.2. Mass spectrometry: molecular-weight determination

Mass spectra were acquired using an ABi 4800 MALDI-TOF/TOF analyzer (Applied Biosystems, Framingham, Massachusetts, USA) in linear mode with an Nd:YAG 355 nm wavelength laser, averaging 1000 laser shots.



**Figure 1**

(a) The best N-PGK crystals were grown using the hanging-drop vapour-diffusion method and a reservoir solution consisting of 32% (w/v) polyethylene glycol 4000 in 0.1 M Tris-HCl pH 7.5 and 0.2 M Li<sub>2</sub>SO<sub>4</sub>. (b) Diffraction image of an N-PGK crystal diffracting to 1.86 Å resolution (1.0° oscillation range). (c) Enlargement showing the highest resolution area.

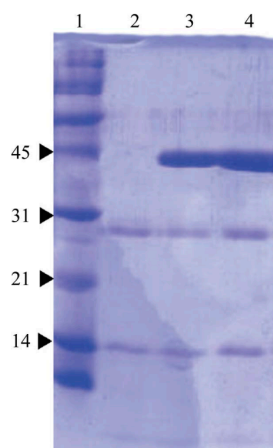


### 2.3. MALDI peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS)

For MALDI-TOF/TOF analysis, data were automatically acquired using an ABI 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Framingham, Massachusetts, USA) in positive-ion reflector mode (the ion-acceleration voltage was 25 kV for MS acquisition and 1 kV for MS/MS) and the obtained spectra were stored in the ABI 4000 Series Explorer Spot Set Manager. PMF and MS/MS fragment-ion spectra were smoothed and corrected to zero baseline using routines embedded in *ABI 4000 Series Explorer* v.3.6 software. Each PMF spectrum was internally calibrated with the mass signals of trypsin-autolysis ions to reach a typical mass measurement accuracy of <25 p.p.m. Known trypsin and keratin mass signals, as well as potential sodium and potassium adducts (+21 and +39 Da), were removed from the peak list. *GPS Explorer* v.4.9 was used to submit the combined PMF and MS/MS data to *MASCOT* software v.2.2.04 (Matrix Science, London, England) by searching in the nonredundant NCBI protein database.

### 2.4. Crystallization

Extensive assays to identify crystallization conditions for PGK were carried out by the sitting-drop vapour-diffusion method at 291 K using Innovaplate SD-2 microplates (Innovadyne Technologies Inc., California, USA). The His tag was retained on the protein during crystallization trials. The initial crystallization screening kits were Crystal Screen HT, Index HT (Hampton Research, Aliso Viejo, California, USA), JCSG+ Suite and PACT Suite (Qiagen, Düsseldorf, Germany). 500 nl droplets were prepared by mixing 250 nl reservoir solution with 250 nl protein solution at a concentration of 5 mg ml<sup>-1</sup> in phosphate-buffered saline (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl adjusted to pH 8.0) with a NanoDrop robot (Innovadyne Technologies Inc.). The droplets were equilibrated against 65 µl reservoir solution. For the cocrystallization trials, the protein solution was prepared as follows. 50 µl protein solution (6 mg ml<sup>-1</sup> in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl adjusted to pH 8.0) was mixed with 5 µl of each of the substrates: 100 mM AMP-PNP (adenylyl imidodiphos-



**Figure 2**  
SDS-PAGE analysis of dissolved PGK crystals. Lane 1, standard broad-range (Bio-Rad) molecular-weight markers (labelled in kDa). Lane 2, dissolved PGK crystals (N-PGK) belonging to space group *I*222; the crystals presented a band of approximately 25 kDa. Lanes 3 and 4, samples of the PGK protein solution stored at 193.15 K for three months. The band at 41.9 kDa corresponds to the full-length protein and the band at 25 kDa corresponds to the N-terminal domain of PGK. A third band is also observed at 14 kDa, which could be related to degradation of the C-terminal domain. Coomassie Blue staining was used for protein detection.

**Table 1**

Data-collection statistics for N-PGK and the ternary complex PGK-AMP-PNP-3PGA.

Values in parentheses are for the highest resolution shell.

	N-terminal PGK domain (N-PGK)	PGK-AMP-PNP-3PGA ternary complex
Crystal data		
Space group	<i>I</i> 222	<i>P</i> 2 <sub>1</sub>
Unit-cell parameters		
<i>a</i> (Å)	62.76	40.35
<i>b</i> (Å)	75.38	78.23
<i>c</i> (Å)	83.63	59.03
$\alpha$ (°)	90	90
$\beta$ (°)	90	96.34
$\gamma$ (°)	90	90
Data collection		
Temperature (K)	100	100
Radiation source	ESRF beamline ID14-1	ESRF beamline ID14-2
Detector	ADSC Quantum Q315r detector	ADSC 210 CCD detector
Wavelength (Å)	0.933	0.934
Resolution range (Å)	31.59–2.08 (2.11–2.08)	46.94–1.78 (1.88–1.78)
No. of images	180	180
Measured reflections	82242	129733
Unique data	12253	34968
Multiplicity	6.7 (6.8)	3.7 (3.6)
Data completeness (%)	100 (100)	99.9 (100)
Average <i>I</i> / $\sigma$ ( <i>I</i> )	29.4 (4.2)	27.5 (2.2)
<i>R</i> <sub>merge</sub> <sup>†</sup>	0.10 (0.42)	0.08 (0.59)

<sup>†</sup> *R*<sub>merge</sub> is defined as  $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where *I*<sub>*i*</sub>(*hkl*) is the intensity of an individual reflection and  $\langle I(hkl) \rangle$  is the average intensity for this reflection; the summation is over all intensities.

phate, an ATP analogue) and 100 mM 3PGA (3-phospho-D-glycerate). All samples were incubated for 1 h on ice with Complete EDTA-free protease inhibitor (Roche Applied Science) just before crystallization assays.

Positive results were scaled up using the hanging-drop vapour-diffusion method in Linbro plates (Hampton Research), mixing 1 µl protein solution with 1 µl reservoir solution and equilibrating the drops against 0.5 ml crystallization solution in the reservoir.

### 2.5. X-ray data collection and processing

X-ray diffraction data sets were collected at 100 K using an ADSC (Area Detector Systems Corporation) Quantum Q315r detector on beamline ID 14-1 ( $\lambda = 0.934$  Å) or an ADSC 210 CCD detector on beamline ID 14-2 ( $\lambda = 0.933$  Å) at the ESRF, Grenoble, France. Cryoprotection of the crystals for data collection was achieved by transferring them for a few seconds into a solution consisting of 15% (v/v) glycerol in the optimized reservoir solution, followed by flash-cooling in liquid nitrogen. Collected data were processed using *iMOSFLM* (Battye *et al.*, 2011) and scaled with *SCALA* (Winn *et al.*, 2011).

### 2.6. Preliminary structure solution

PGK structure determination was initiated using the *Thermotoga maritima* PGK structure (PDB entry 1vpe; Auerbach *et al.*, 1997), which shows 55.4% sequence identity to pneumococcal PGK. This model was used as a template in the molecular-replacement method using *MOLREP* (Vagin & Teplyakov, 2010).

## 3. Results

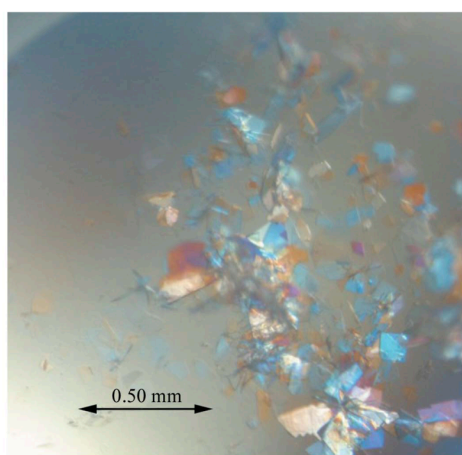
PGK crystals grew within 1 d from 30% (w/v) polyethylene glycol (PEG) 4000, 200 mM Li<sub>2</sub>SO<sub>4</sub> and 100 mM Tris-HCl pH 8.5 (condition No. 17 of Crystal Screen HT; Hampton Research). Optimization of

## crystallization communications

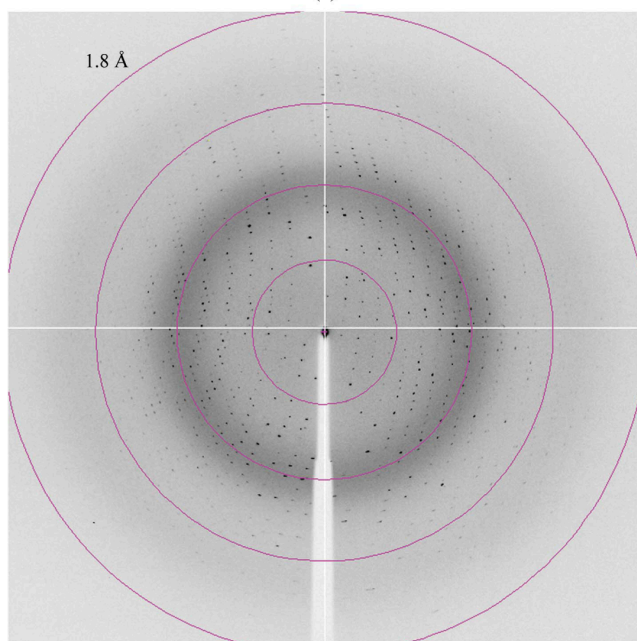
crystal quality was achieved by testing different pH values and by lowering the PEG concentration.

The largest PGK crystals, with maximum dimensions of about  $0.5 \times 0.25 \times 0.5$  mm, were obtained after optimization by the hanging-drop vapour-diffusion method against a reservoir solution consisting of 32% (w/v) PEG 4000, 200 mM  $\text{Li}_2\text{SO}_4$  and 100 mM Tris-HCl pH 7.5. (Fig. 1a).

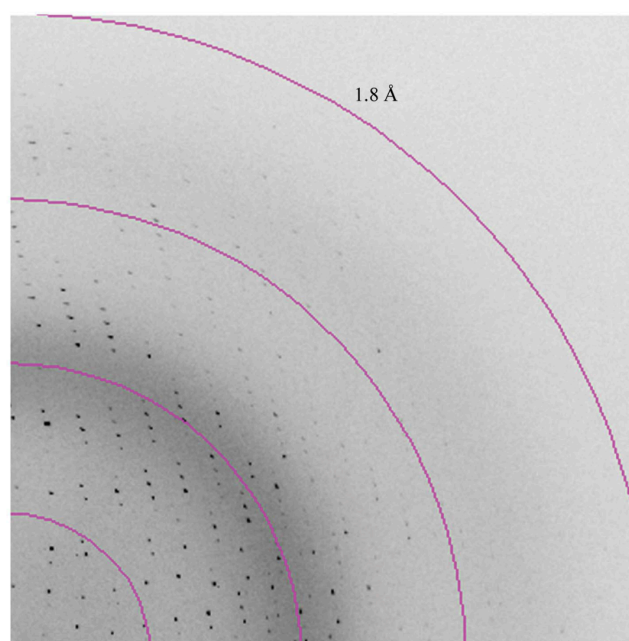
An X-ray data set was collected to 2.0 Å resolution from a PGK crystal and showed a well defined diffraction pattern (Fig. 1b). The crystals belonged to the orthogonal space group  $I222$ , with unit-cell parameters  $a = 62.76$ ,  $b = 75.38$ ,  $c = 83.63$  Å. Specific volume calculations based on the molecular weight of full-length PGK (Matthews, 1968) indicated that the asymmetric unit could not even contain one PGK molecule (solvent content of 4.34% and  $V_M = 1.18 \text{ Å}^3 \text{ Da}^{-1}$ ).



(a)



(b)



(c)

**Figure 3**

(a) Crystals of the PGK-AMP-PNP-3PGA ternary complex were obtained using the hanging-drop vapour-diffusion method and a reservoir solution consisting of 30% (w/v) polyethylene glycol 4000, 0.2 M  $\text{MgCl}_2$  and 0.1 M Tris-HCl pH 8.6. The drops were prepared by mixing 2  $\mu\text{l}$  reservoir solution with 4  $\mu\text{l}$  protein solution (5 mg  $\text{ml}^{-1}$ ) and including 0.5  $\mu\text{l}$  25 mM *n*-decanoyl sucrose as an additive. (b) Diffraction image of a PGK-AMP-PNP-3PGA crystal diffracting to 1.78 Å resolution ( $1.0^\circ$  oscillation range). (c) Enlargement showing the highest resolution area.

The obtained solvent content was obviously not compatible with the presence of full-length PGK in the crystals.

In order to understand this result, we decided to carry out SDS-PAGE analysis, a mass-spectrometric assay and fingerprinting of the PGK crystals. The SDS-PAGE analysis (Fig. 2) indicated that the protein in the crystal was proteolyzed, presenting a molecular mass of 25 kDa instead of the 41.9 kDa expected for the full-length protein. Mass spectrometry confirmed these results. Fingerprinting of the PGK crystals indicated the presence of only the N-terminal domain of PGK (N-PGK) in the crystals (220 amino acids). The Matthews coefficient was recalculated assuming one N-PGK molecule per asymmetric unit, yielding a  $V_M$  of  $2.39 \text{ Å}^3 \text{ Da}^{-1}$  and a solvent content of 48.63%.

In order to gain stability and avoid proteolysis of the full-length PGK, crystallization in the presence of AMP-PNP and 3PGA was attempted. Crystals of the ternary PGK-AMP-PNP-3PGA complex grew from a reservoir solution consisting of 30% (w/v) PEG 4000, 0.2 M  $\text{MgCl}_2$  and 0.1 M Tris-HCl pH 8.6. Drops were prepared by mixing 2  $\mu\text{l}$  of this reservoir solution with 4  $\mu\text{l}$  protein solution (5 mg  $\text{ml}^{-1}$  in the same buffer as previously) and including 0.5  $\mu\text{l}$  25 mM *n*-decanoyl sucrose as an additive (Fig. 3a). Crystals of the ternary PGK-AMP-PNP-3PGA complex presented a different habit to that of the native crystals. The ternary complex crystals grew to maximum dimensions of  $0.25 \times 0.10 \times 0.25$  mm within 3 d at a temperature of 291.15 K. An X-ray data set for the ternary PGK-AMP-PNP-3PGA complex was collected to 1.78 Å resolution and displayed a well defined diffraction pattern (Fig. 3b). X-ray diffraction data were processed and showed good data-collection statistics (Table 1). These crystals belonged to the monoclinic space group  $P2_1$ , with unit-cell parameters  $a = 40.35$ ,  $b = 78.23$ ,  $c = 59.03$  Å,  $\beta = 96.35^\circ$ . Specific volume calculations based on the molecular mass of full-length PGK (398 amino acids, 41 953 Da) indicated the probable

## crystallization communications

presence of one PGK molecule per asymmetric unit, with 44.27% solvent content and a Matthews coefficient  $V_M$  of  $2.21 \text{ \AA}^3 \text{ Da}^{-1}$ . Molecular replacement further confirmed the presence of full-length PGK protein in the crystals of the ternary complex. These results revealed that the presence of AMP-PNP and 3PGA prevented the proteolysis observed for the N-PGK crystals. Structural refinement of the ternary PGK–AMP–PNP–3PGA complex is currently in progress.

## References

- Auerbach, G., Huber, R., Grättinger, M., Zaiss, K., Schurig, H., Jaenicke, R. & Jacob, U. (1997). *Structure*, **5**, 1475–1483.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst. D* **67**, 271–281.
- Bergmann, S. & Hammerschmidt, S. (2006). *Microbiology*, **152**, 295–303.
- Bergmann, S., Rohde, M., Preissner, K. T. & Hammerschmidt, S. (2005). *Thromb. Haemost.* **94**, 304–311.
- Bergmann, S., Wild, D., Diekmann, O., Frank, R., Bracht, D., Chhatwal, G. S. & Hammerschmidt, S. (2003). *Mol. Microbiol.* **49**, 411–423.
- Bernstein, B. E., Michels, P. A. & Hol, W. G. (1997). *Nature (London)*, **385**, 275–278.
- Hammerschmidt, S. (2006). *Curr. Opin. Microbiol.* **9**, 12–20.
- Hughes, M. J. *et al.* (2002). *Infect. Immun.* **70**, 1254–1259.
- Jin, H., Song, Y. P., Boel, G., Kochar, J. & Pancholi, V. (2005). *J. Mol. Biol.* **350**, 27–41.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Pancholi, V., Fontan, P. & Jin, H. (2003). *Microb. Pathog.* **35**, 293–303.
- Pecorari, F., Guilbert, C., Minard, P., Desmadril, M. & Yon, J. M. (1996). *Biochemistry*, **35**, 3465–3476.
- Terao, Y., Yamaguchi, M., Hamada, S. & Kawabata, S. (2006). *J. Biol. Chem.* **281**, 14215–14223.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.



#### 4.1.1. DISCUSIÓN

La proteína PGK se expresó y purificó con cola de histidinas en el tampón regulador 50 mM  $\text{NaH}_2\text{PO}_4$  pH 8,0 y 300 mM NaCl.

Se ensayaron más de 700 condiciones diferentes de cristalización a través de baterías soluciones comerciales, en una de estas soluciones se obtuvieron cristales iniciales. La proteína PGK cristaliza en un día a una concentración de  $5 \text{ mg mL}^{-1}$  en la condición de cristalización: 30% (m/v) de polietilenglicol (PEG) 4000, 200 mM  $\text{Li}_2\text{SO}_4$  y 100 mM Tris-HCl a pH 8,5.

Los primeros cristales obtenidos se midieron en una fuente de radiación sincrotrón, en el ESRF (Grenoble, Francia). Se obtuvo un conjunto de datos a  $2,0 \text{ \AA}$  de resolución, los cristales pertenecían al grupo espacial ortorrómbico  $I222$ , con los parámetros de celdilla  $a = 62,76$ ,  $b = 75,38$ ,  $c = 83,63 \text{ \AA}$ . Los volúmenes de celdilla no permitían contener la proteína completa por lo que se determinó a través de un gel de electroforesis y espectrometría de masas que la proteína estaba proteolizada.

Para evitar la proteólisis se incubó la proteína en presencia de los sustratos de la reacción, AMP-PNP y 3PG. Se obtuvieron así cristales del complejo PGK-AMP-PNP-3PG en la condición de cristalización: 30% (m/v) PEG 4000, 200 mM  $\text{MgCl}_2$  y 100 mM Tris-HCl a pH 8,6. La optimización se llevo a cabo mediante el uso del detergente *n*-decanoyl sucrose (25 mM).

Los cristales del complejo crecen en 3 días con un hábito cristalino diferente al que presentaban los de la proteína sin ligandos. Se recogió un conjunto de datos de difracción en el sincrotrón ESRF (Grenoble, Francia) a  $1,78 \text{ \AA}$  de resolución, con buenas estadísticas.

Los cristales del complejo ternario pertenecen al grupo espacial monoclinico  $P2_1$ , con los parámetros de celdilla  $a = 40,35$ ,  $b = 18,23$ ,  $c = 59,03 \text{ \AA}$ ,  $\beta = 96,35^\circ$ , con una molécula por unidad asimétrica.

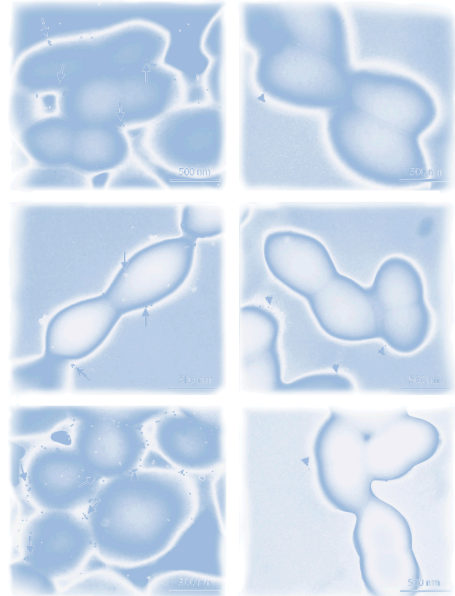
Mediante este trabajo de investigación se pueden obtener las siguientes conclusiones:

- ✓ Se ha expresado y purificado la proteína PGK.
- ✓ Se han obtenido las condiciones óptimas de cristalización de la proteína PGK.
- ✓ Se ha establecido un protocolo para evitar la proteólisis de la PGK.
- ✓ Se ha conocido el grupo espacial de la PGK, así como los parámetros de celdilla.
- ✓ Se ha recogido una conjunto de datos apropiados para la determinación estructural de la proteína.





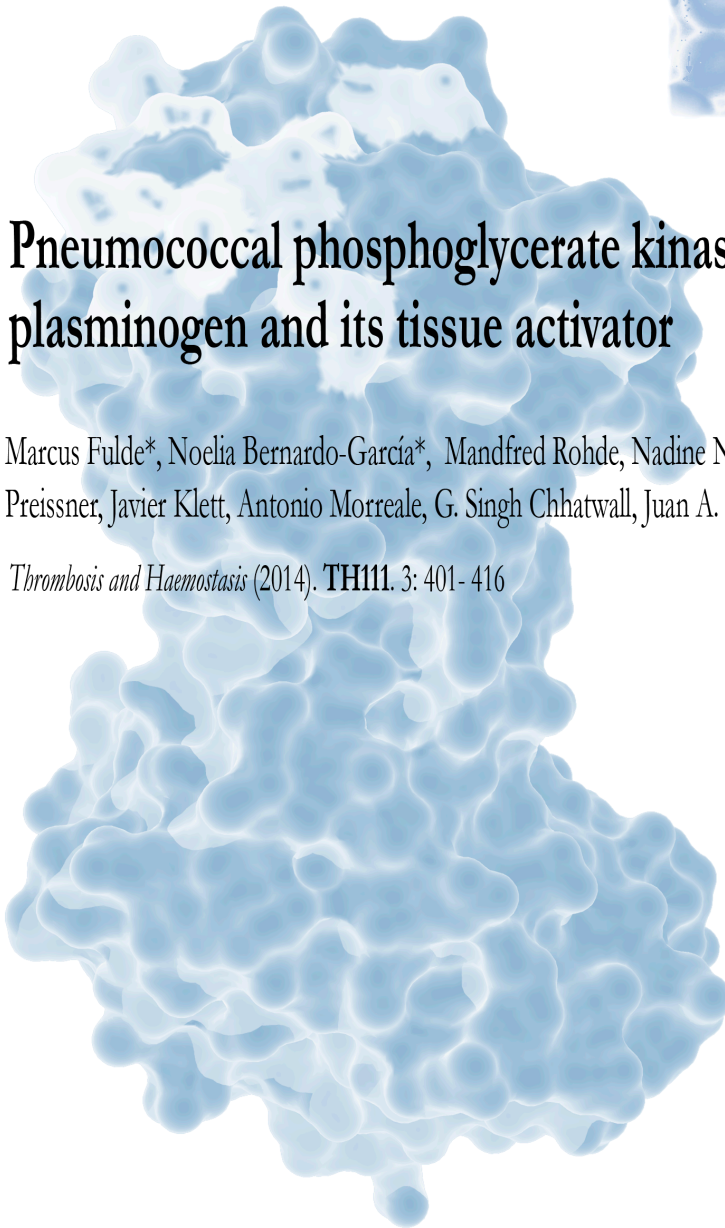
## Trabajo de Investigación II:



### Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator

Marcus Fulde\*, Noelia Bernardo-García\*, Manfred Rohde, Nadine Natchigall, Ronald Frank, Klauss T. Preissner, Javier Klett, Antonio Morreale, G. Singh Chhatwall, Juan A. Hermoso and Simone Bergmann

*Thrombosis and Haemostasis* (2014). **TH111**. 3: 401- 416







## 4.2. INTRODUCCIÓN A LA FOSFOGLICERATO QUINASA Y APORTACIÓN DEL AUTOR

*Streptococcus pneumoniae* es el causante de una gran variedad de enfermedades infecciosas, coloniza el tracto respiratorio superior de humanos y causa infecciones locales como la otitis. Además, *S. pneumoniae* es conocido por ser el responsable de infecciones severas como la neumonía y la meningitis en niños, pacientes de edad avanzada y pacientes inmunodeprimidos. El aumento de la resistencia a antibióticos por parte del patógeno hace necesaria la búsqueda nuevos medios para combatirlo.

Neumococo presenta en su superficie celular diversas moléculas que interaccionan con el hospedador a través de receptores, proteínas de la matriz extracelular y componentes del sistema circulatorio. Así, la unión al plasminógeno, (PLG) el componente mayoritario en el proceso de la fibrinólisis, favorece la diseminación y dispersión del patógeno a través de los tejidos y del sistema vascular.

Se han identificado una serie de enzimas metabólicas, como la enolasa y la GADPH, en la superficie del patógeno, denominándolas proteínas de superficie no clásicas (NCSP). Estas proteínas exhiben funciones variadas, incluyendo la adherencia a componentes del hospedador como la fibronectina, laminina y receptores para el reclutamiento del PLG. La fosfoglicerato kinasa (PGK) de *S. agalactiae* también ha sido clasificada como una proteína que presenta funciones adicionales a la que realiza en la glicólisis.

A través de este trabajo nos planteamos los siguientes objetivos:

- ✓ Resolver la estructura tridimensional de la proteína PGK.
- ✓ Identificar a la PGK como una proteína no clásica de superficie del neumoco y conocer la localización de la PGK en la superficie celular del neumococo.
- ✓ Caracterizar el sitio de interacción de la PGK con el plasminógeno.
- ✓ Identificar el mecanismo de interacción de la PGK con el plasminógeno.
- ✓ Proponer, por primera vez, un mecanismo de interacción de una proteína de superficie con un activador del plasminógeno. Este mecanismo fue comprobado experimentalmente.

En esta publicación, la autora de esta Tesis ha realizado la resolución de la estructura tridimensional de la PGK, así como del análisis estructural, ha identificado el otro sitio de unión al PLG y propuesto su interacción con el activador de PLG. También ha propuesto un modo de interacción de la proteína PGK con proteínas del hospedador.



**Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator.**

Marcus Fulde\*, Noelia Bernardo-García\*, Manfred Rohde, Nadine Nachtigall, Ronald Frank, Klauss T. Preissner, Javier Klett, Antonio Morreale, G. Singh Chhatwall, Juan A. Hermoso y Simone Bergmann

Thrombosis and Haemostasis (2014) TH**111**. 3: 401- 416



## Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator

Marcus Fulde<sup>1,2\*</sup>; Noelia Bernardo-García<sup>3\*</sup>; Manfred Rohde<sup>1</sup>; Nadine Nachtigall<sup>1</sup>; Ronald Frank<sup>4</sup>; Klaus T. Preissner<sup>5</sup>; Javier Klett<sup>6</sup>; Antonio Morreale<sup>6,7</sup>; G. Singh Chhatwal<sup>1</sup>; Juan A. Hermoso<sup>3</sup>; Simone Bergmann<sup>1,8</sup>

<sup>1</sup>Department of Medical Microbiology, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany; <sup>2</sup>Current address: Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany; <sup>3</sup>Department of Crystallography and Structural Biology, Instituto de Química-Física "Rocasolano", CSIC, Madrid, Spain; <sup>4</sup>Department of Chemical Biology, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany; <sup>5</sup>Institute for Biochemistry, Justus-Liebig University of Giessen, Giessen, Germany; <sup>6</sup>Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Campus UAM, Madrid, Spain; <sup>7</sup>Current address: Repsol, Technology Center, Móstoles, Madrid, Spain; <sup>8</sup>Current address: Department of Infection Biology, Institute of Microbiology, Technische Universität Braunschweig, Braunschweig, Germany

### Summary

*Streptococcus pneumoniae* is not only a commensal of the nasopharyngeal epithelium, but may also cause life-threatening diseases. Immune-electron microscopy studies revealed that the bacterial glycolytic enzyme, phosphoglycerate kinase (PGK), is localised on the pneumococcal surface of both capsulated and non-capsulated strains and colocalises with plasminogen. Since pneumococci may concentrate host plasminogen (PLG) together with its activators on the bacterial cell surface to facilitate the formation of plasmin, the involvement of PGK in this process was studied. Specific binding of human or murine PLG to strain-independent PGK was documented, and surface plasmon resonance analyses indicated a high affinity interaction with the

kringle domains 1–4 of PLG. Crystal structure determination of pneumococcal PGK together with peptide array analysis revealed localisation of PLG-binding site in the N-terminal region and provided structural motifs for the interaction with PLG. Based on structural analysis data, a potential interaction of PGK with tissue plasminogen activator (tPA) was proposed and experimentally confirmed by binding studies, plasmin activity assays and thrombus degradation analyses.

### Keywords

Angiostatin, plasminogen, phosphoglycerate kinase, *Streptococcus pneumoniae*, tissue-type plasminogen activator

### Correspondence to:

Simone Bergmann, PhD  
Technische Universität Braunschweig  
Spielmannstrasse 7, 38106 Braunschweig, Germany  
Tel.: +49 531 391 5818, Fax: +49 531 391 5854  
E-mail: simone.bergmann@tu-bs.de

### Financial support:

The research leading to these results has received funding from the European Community's Seventh Framework Program under Grant Agreement no. HEALTH-F3-2009-223111. This work was also supported by grants from the Spanish Ministry of Economy and Competitiveness (BFU2011-25326) and Comunidad Autónoma de Madrid (CAM) S2010-BMD-2457 (BIPEDD2). J.K. is funded by a grant from Ministerio de Economía y Competitividad (BFU2011-24595). A.M. also acknowledges CAM for financial support to the Fundación Severo Ochoa through the AMAROUTO program.

\* These authors contributed equally to the work.

Received: May 24, 2013

Accepted after major revision: October 1, 2013

Prepublished online: November 7, 2013

doi:10.1160/TH13-05-0421

Thromb Haemost 2014; 111: 401–416

## Introduction

Infectious diseases caused by *Streptococcus pneumoniae* (*pneumococcus*) represent a major and increasing life-threatening fact in both industrialised and developing countries. As commensals, pneumococci colonise the upper respiratory tract of humans and mediate local infections, e.g. otitis media. More important, *S. pneumoniae* is known to cause severe systemic diseases like pneumonia and meningitis in children, elderly patients and immune compromised people (1). Increasing antibiotic resistances state reason for the current intensive efforts in vaccine-based prevention strategies and lead to maintenance of high mortality and morbidity rates of pneumococcal infections (1, 2). Pneumococci express different surface molecules mediating interactions with host

cell receptors, extracellular matrix proteins, and components circulating in the vascular system (3, 4). Binding and activation of the major fibrinolysis protein plasminogen (PLG) has been extensively investigated during the last decade. Its recruitment on the bacterial surface provides an effective precaution to subvert the activity of the serine protease plasmin. As proteolytic cofactor, surface-bound plasmin enables pneumococci to spread intercellularly and disseminate in deeper tissue sites and within the vascular system (5, 6).

PLG is a 92 kDa pro-enzyme of the broad-spectrum protease plasmin, which counteracts the blood coagulation cascade. It is composed of a pre-activation peptide (~8 kDa), five homologous disulfide-bonded triple-loop kringle structures (K1–5, 65 kDa) and a serine-protease domain (7). The first four kringle domains

of PLG act as an inhibitory molecule for angiogenesis and are, therefore, termed angiostatin (8), whereas the last kringle domain together with the enzymatic domain is known as mini-plasminogen (mini-PLG) (9). PLG is converted into plasmin by eukaryotic plasminogen activators (urokinase-type plasminogen activator, uPA and tissue-type plasminogen activator, tPA) (10). These activators initiate the conversion of PLG to plasmin, which hydrolyses polymerised fibrin strands into soluble fibrin degradation products. Increased levels of tPA, urokinase-type plasminogen activator receptor (uPAR), plasminogen activator inhibitors 1 and 2 (PAI-1, PAI-2), and also uPA have been reported in the cerebrospinal fluids of patients suffering from pneumococcal meningitis (11).

A variety of bacterial metabolic enzymes, such as enolase and glyceraldehyde-3-phospho dehydrogenase (GAPDH), have been characterised as surface located proteins, called non-classical surface proteins (4). These proteins exhibit "moonlighting" functions including adherence to various host components like fibronectin (12) and laminin (13) and as receptors for PLG recruitment (6, 14). The phosphoglycerate kinase (PGK) was identified as an additional glycolytic enzyme bearing moonlighting function on the surface of *S. agalactiae* (15). The mechanism of cell wall secretion and the mode of surface anchorage remains an unsolved question. The pneumococcal PGK represents a monomeric two-domain protein of 41.9 kDa that mediates transfer of anhydride phosphate of 1,3-biphospho-D-glycerate (1,3-BPG) to Adenosin 5'-diphosphate magnesium salt (MgADP), yielding 3-phospho-D-glycerate (3-PGA) and Adenosin 5'-triphosphate magnesium salt (MgATP) within glycolytic metabolism (16). The N-terminal domain contains the substrate binding site followed by an ATP-binding site and the catalytic active centre in the C-terminus (17). The present study provides insights in the moonlighting functions of pneumococcal PGK by biophysical and structural analyses.

## Material and methods

### Proteins and antibodies

Polyclonal antibodies recognising recombinant pneumococcal PGK were raised in rabbits and were purchased commercially (Pineda, Berlin, Germany). Human PLG was purchased from Sigma-Aldrich (St. Louis, MO, USA), tPA, uPA and thrombin (derived from bovine plasma) were purchased from Haemochrom Diagnostica (Essen, Germany). Protein fractions containing kringle domains 1–3 or 1–4 (angiostatin) from PLG, mini-plasminogen, and murine plasminogen were gratefully obtained from K.T. Preissner (University of Giessen, Germany). Anti-PLG antibody from goat was purchased from Affinity Biologicals (GAPG-IG; Affinity Biologicals, Ancaster, ON, Canada), HRP-conjugated polyclonal rabbit anti goat antibody from Dako (PO449, 0.5 g/l; Dako, Glostrup, Denmark).

### Bacterial strains and culture

*Streptococcus pneumoniae* serotype 2 strain D39, corresponding capsule-deficient mutant (D39Δcps; [18]) and serotype 35A pneumococci (ATCC11733) were used for electron microscopic studies and amplification of the *pgk*-gene for expression cloning. Pneumococcal strains were cultured at 37°C in Todd-Hewitt broth (Oxoid, Hampshire, UK) supplemented with 0.5% yeast extract (THY) to mid log phase or were grown on blood agar plates (Becton Dickinson, Heidelberg, Germany). *Escherichia coli* M15[pREP4] (Qiagen, Hilden, Germany) was used as host strain for recombinant pQE expression plasmids (Qiagen) and was cultured at 37°C on Luria-Bertani (LB) agar or grown on LB agar containing 100 µg/ml ampicillin and 50 µg/ml kanamycin.

### Cloning, expression and protein purification of recombinant PGK

Expression and purification of PGK was performed as described in (19). In brief, the *pgk* gene from *S. pneumoniae* D39 (Acc. No: YP\_815953.1) was amplified with oligonucleotides (5'-GGGGGATCCTTGGCAAACTTACTGTATAAGAG-3') and (5'-CCCGTCGACTTATTTTCTGTCAAGGCTGCAAG-3') (MWF Biotech, Ebersberg, Germany) and ligated into the pQE30 plasmid (Qiagen) generating pQE30PGKwt in *E. coli* M15. After induction of protein expression with 1 mM IPTG, *E. coli* cells were lysed mechanically using a French Pressure Cell (SLM-Aminco). Purification of recombinant PGK was performed by affinity chromatography according to standard protocols recommended by Qiagen.

### Protein dot spot analysis

Recombinant PGK proteins, purified human or murine PLG, PLG derivatives and tissue-type plasminogen activator (tPA) were immobilised onto a nitrocellulose membrane. After blocking with 10% skimmed milk, the membrane was incubated with 40 µg of PLG or 40 µg PGK protein for 4 hours (h) at 4°C. After three wash steps with PBS (phosphate-buffered saline, 80 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 2.7 mM KCl, 137 mM NaCl, pH 7.4), supplemented with 0.05% Tween 20, the membrane was incubated for 1 h at room temperature with PLG-specific antibody from goat or PGK-specific antibody raised in rabbit followed by incubation with horse radish peroxidase-conjugated secondary antibodies recognising goat or rabbit for 1 h at room temperature. Signal detection was performed with 1 mg/ml 4-chloro-1-naphthol and 0.1% H<sub>2</sub>O<sub>2</sub> in PBS or enhanced chemiluminescence (Pierce, Rockford, IL, USA).

### Surface plasmon resonance analyses

The association and dissociation reactions of human and murine PLG to recombinant PGK were analysed in the BIAcore optical biosensor (BIAcore 2000 system, GE Healthcare, Freiburg, Germany) using CM5 sensor chips. Covalent immobilisation of PLG

was performed using a standard amine coupling procedure essentially described previously (20). Binding analyses was performed in HBS BIAcore-running buffer (10 mM Hepes, 150 mM NaCl, and 0.05 % Tween 20 (pH 7.4) without EDTA at 20°C using a flow rate of 30 µl/minute (min). PGK was used as analyte in concentrations of 4.0 µM, 2.0 µM, 1.0 µM and 0.5 µM. The affinity surface was regenerated with 10 µl of 20 mM NaOH. Binding was assayed at least in duplicates using independently prepared sensor chips. Parameters of binding kinetics were analysed from raw data of the BIAcore sensograms suitable for analysis using the kinetic models included in the BIAevaluation software version 3-0. The experimental data were fitted globally using a 1:1 Langmuir kinetic). For each evaluation a minimum of four data sets were analysed.

### Peptide spot array

The peptide spot array was performed essentially as described (19, 20). The highly conserved 398 amino acid sequence of the pneumococcal PGK was subdivided into 16mers with offset of four residues. Based on the sequences of three peptide spots, which were recognised to be positive for PLG binding, a systemic mutational analysis was performed. Peptides were synthesised containing alanine amino acid exchanges at each position of the original peptide sequence. For assaying PLG binding, membranes were treated as described previously (20). For protein overlay the membrane was incubated with 50 µg human PLG for 4 h at room temperature. Detection of binding was performed with PLG-specific antibodies and horse radish peroxidase conjugated secondary antibody for 1 h each at room temperature. For visualisation of binding signals, the membrane was subjected to enhanced chemiluminescence reaction (Super Signal West Pico, Thermo Scientific, Ulm, Germany) following the instructions of the manufacturer. In order to exclude unspecific binding signals, the membrane was treated equally without the step of PLG incubation. A specific washing procedure based on a stepwise incubation with dest. H<sub>2</sub>O, N,N-dimethylformamid and 8 M urea-containing buffer allowed a repetition of binding studies.

### PGK crystallisation, data collection and processing

Crystals for X-ray analysis were prepared as previously described (21) using 30% PEG 4000, 0.2 M MgCl<sub>2</sub> and Tris-HCl at pH 8.6 in the presence of AMP-PNP (adenylyl-imidodiphosphate) (100 mM) and 3-PGA (3-phospho-D-glycerate) (100 mM). Crystals belonged to the monoclinic space group P2<sub>1</sub> with unit-cell parameters  $a = 40.35$  Å,  $b = 78.23$  Å,  $c = 59.03$  Å and  $\beta = 96.35^\circ$ . X-Ray diffraction data sets were collected using an ADSC210 CCD detector ( $\lambda = 0.933$  Å) on beamline ID 14-2 at 100 K in the ESRF facility of Grenoble. Collected data from cryo-protected crystals were processed using IMOSFLM (22) and scaled with SCALA, (Collaborative Computation Project, 1994).

### PGK structure determination and refinement

PGK structure was solved by the molecular replacement method using the *MOLREP* program from the CCP4 package with the *Thermotoga maritima* PGK structure (PDB code 1VPE) as the initial model (23). Electron density maps for model building were calculated using conventional 2Fo-Fc and Fo-Fc coefficients with PHENIX suite (24). Intensive model building was performed with COOT (25). Excellent density maps were obtained for the PGK structure except for the two last residues of the polypeptide chain. Electron density corresponding in shape and size to AMP-PNP was observed in maps calculated in the final stages of model building and refinement. Four glycerol molecules used as cryoprotectant were found on the protein molecular surface involved in crystal packing interactions. The refinement converged to the final values of  $R = 0.14$  and  $R_{\text{free}} = 0.19$  at 1.78 Å resolution with 404 waters. Structural validation was done with a Ramachandran plot from PHENIX, it shows that over 98% of the non-glycine, non-proline residues are found in the most favored regions, with no residues in disallowed regions.

### Modelling of PGK in complex with human kringle structures

Initial structures of human Kringle 2 (PDB code 1B2I), Angiostatin (PDB code 1KI0), human tissue-type plasminogen activator (tPA, PDB code 1TPK) and urokinase plasminogen activator (uPA, PDB code 2I9A) were manually oriented facing their Kringle binding sites to the PGK PLG binding site and uploaded to Rosseta Dock Server (RDS) (26). A total of 1,000 solutions were kept from each docking and filtered to select those containing the interactions involving Lys14, Lys15 and Arg53 residues. For each pair of structures a contact-overlap, based on counting common contacts, was calculated and arranged in a similarity matrix. The Average Linkage clustering algorithm was applied to the matrix using the standard Euclidean distance as a classification criterion and a maximum of 30 clusters were retained (*MATLAB, The MathWorks Inc.*, 7.11.0; Natick, MA, USA, 2010). The most stable pose according to its ISM binding energy (see below) within the most populated cluster was selected as the docking pose.

Selected docking poses were minimised in vacuum with the AMBER 03 force field (27, 28) during 10,000 steps (the first 1,000 with the steepest descent method and the rest with the conjugated gradient) solely to alleviate the possible clashes that may be left after the docking program. All calculations were done with AMBER 11 package (29). The resulting structures were then processed with ISM (30). In brief, ISM calculates the binding energy according to the following equation:

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} + \Delta G_{\text{desolv}}^{\text{R}} + \Delta G_{\text{desolv}}^{\text{L}} + \Delta G_{\text{apo}} + \Delta G_{\text{HB}}$$

where  $\Delta G_{\text{binding}}$  represents van der Waals interactions calculated through the 12-6 Lennard-Jones potential;  $\Delta G_{\text{elec}}$  accounts for charge-charge electrostatic interactions modelled with a sigmoidal dielectric screening function;  $\Delta G_{\text{desolv}}^{\text{R}}$  and  $\Delta G_{\text{desolv}}^{\text{L}}$  are the desol-

valuations terms corresponding to receptor and ligand, respectively;  $\Delta G_{np}$  is a solvent accessible-related term to account for shape complementarity and the loss in surface area produced upon complex formation; and  $\Delta G_{HB}$  is the hydrogen bonding term. Residue-residue interactions were classified into three types: a) hydrogen bonding; b) hydrophobic (which includes  $\Delta G_{vdw}$  and non-polar contribution from  $\Delta G_{desolv}^R$ ,  $\Delta G_{desolv}^L$ , and  $\Delta G_{apo}$ ; and c) hydrophilic (which includes  $\Delta G_{elec}$  and the polar contribution from  $\Delta G_{desolv}^R$ ,  $\Delta G_{desolv}^L$ , and  $\Delta G_{apo}$ ). Molecular visualisation was performed with PyMOL program (DeLano, 2002).

### Electron microscopic visualisation

Pre-embedding labelling of *S. pneumoniae* serotype 35A and 2 D39 wild-types and capsule-deficient mutant was performed as described before (14) using 100 µg/ml PGK-specific IgG antibody (protein A-sepharose purified) and a 1:10 dilution of the stock solution of 15 nm protein A gold-nanoparticles. After fixation in 1% glutaraldehyde bacteria were adsorbed onto carbon-coated Formvar grids, washed in TE buffer and distilled water before air drying. Samples were examined in a Zeiss EM910 transmission electron microscope at calibrated magnifications and at an acceleration voltage of 80 kV. Images were recorded digitally with a Slow-Scan CCD-Camera (ProScan, 1024x1024, Scheuring, Germany) with ITEM-Software (Olympus Soft Imaging Solutions, Münster, Germany).

For colocalisation studies, pre-embedding labelling of surface-displayed PGK and recruited plasminogen was performed with capsule-deficient mutant of *S. pneumoniae* D39. After fixation of bacteria with 1% formaldehyde directly in THY culture medium, bacteria were harvested by centrifugation (5,000 rpm, 10 min). For preincubation with PGK,  $5 \times 10^9$  bacteria were incubated for 30 min at room temperature with 100 µg purified PGK protein followed by three washing steps with PBS. PGK-preincubated and non-preincubated bacteria were washed in PBS containing 10 mM glycine for quenching of free aldehydes. After washing in PBS, bacteria were resuspended in 200 µl of plasminogen-gold nanoparticles (15 nm in size), incubated for 30 min at 30°C and washed twice with PBS. Then, bacteria were resuspended in 200 µl of a 1:25 solution in 100 mM PBS of rabbit anti-PGK antibodies (1.8 mg/ml IgG, generated by rabbit immunisation with recombinant *S. pneumoniae* PGK), incubated for 30 min at 30°C and washed twice with 100 mM PBS. Bound antibodies were detected with protein A/G gold nanoparticles (10 nm in size) by incubation for 30 min at 30°C. As control bacteria were incubated with 1:10 dilution of PGK-preimmune serum followed by incubation with 15 nm protein A/G gold particles. After washing twice in 100 mM PBS bacteria were fixed with 2% glutaraldehyde at room temperature for 15 min and washed in PBS and TE (Tris 20 mM, 1 mM EDTA, pH 6.9). After resuspension in TE-buffer, bacteria were placed onto butvar-coated nickel grids (300 mesh), left for 10 min to absorb to the plastic film, washed in distilled water and air dried. Samples were observed in a Zeiss Merlin field emission scanning electron microscope (FESEM) at an acceleration voltage of 15 kV. Imaging was performed with the high-efficiency SE2

Everhart-Thornley detector. Contrast and brightness of EM pictures were adjusted using the software program Adobe Photoshop CS5.

### Immobilisation of recombinant PGK to Dynabeads®

Covalent immobilisation of recombinant PGK, human PLG and BSA (bovine serum albumin, Carl Roth GmbH, Karlsruhe, Germany) to carboxylated polystyrene Dynabeads® (My One™ Carboxylated acid, 1.0 µm in size; Invitrogen, Carlsbad, CA, USA) was performed as described by the manufacturer. In brief 280 µg recombinant PGK protein or PLG or BSA was incubated with 660 µl beads ( $\sim 6 \times 10^9$  beads) over night at 4°C, followed by blocking with BSA (Roth). Protein coupling was visualised by SDS-Page.

### Plasmin activity assay

Plasmin activity was determined as described in Fulde et al. (51) using the chromogenic substrate D-valyl-leucyl-lysine-p-nitroanilide dihydrochloride (S-2251, Fluka). In brief, 10 µl of PGK-coated Dynabeads® ( $\sim 5 \times 10^8$  beads) were incubated for 30 min with 50 µg PLG. After three washing steps with PBS using a magnet, the beads were incubated with 600 ng tPA or 600 ng uPA. In parallel, 5.0 µg of PGK-coated Dynabeads® were incubated with 10 µg of tPA or uPA for 30 min. After three washing steps with PBS, the beads were incubated with 3.0 µg PLG. Plasmin activity was determined in time intervals of 5.0 min up to 60 min in triplicates using the Varioscans Flash ELISA Reader (Thermo Scientific) at 405 nm. Measurement started immediately after adding 30 µl of plasmin substrate S-2251 (Haemochrom Diagnostica). For evaluation of plasmin activity, background values of all samples without activator or PLG was determined and mean values was subtracted from tPA-, uPA-, and PLG- incubated samples at each time point of measurement. BSA-coated beads were applied with and without PLG incubation and activator in order to determine background plasmin activity. PLG-coated beads were incubated with tPA or uPA as positive controls.

### Degradation of fibrin matrix

Solid fibrin matrix was produced on transwell insets (falcon, membrane with 3 µm pores) for photographic visualisation and on cover slips for electron microscopy as described in Fulde et al. (52). In brief, 100 µl of 50 mg/ml plasminogen-depleted fibrinogen (Haemochrom Diagnostica) was incubated with 3.0 µl of 1.0 KU/ml thrombin over night at 37°C. After PLG incubation,  $\sim 2.5 \times 10^7$  PGK coated Dynabeads® were incubated with aggregated fibrin matrix and plasmin formation was induced by adding 600 ng tPA or uPA. To exclude unspecific background degradation, BSA-coated Dynabeads® were incubated with plasminogen as indicated above and incubated with the aggregated fibrin matrix with and without 600 ng tPA or uPA. As positive control, PLG-coated Dynabeads® were incubated with fibrin matrix together with 600 ng tPA. Dissolution of fibrin was determined by eye and documented photographically in a time series of two hours for up to 24 h.



Electron microscopic evaluation of dissolved fibrin matrices was performed as described in Fulde et al. (52). In brief, dissolution of fibrin matrices by plasminogen-pretreated PGK-coated beads was initiated by adding 600 ng tPA as plasminogen activator and incubated for 3 h at 37°C. Dissolution was stopped by adding 2% glutaraldehyde in PBS. As control, fibrin thrombi were incubated with PLG-preincubated PGK-coated beads without adding a plasminogen activator. BSA-coated Dynabeads® were used as further control after preincubation with PLG either with or without tPA. To visualise the intact fibrin matrix, aggregated fibrin thrombi were incubated for 3 h with PBS alone. After fixation, samples were washed with TE-buffer (20 mM TRIS, 1 mM EDTA, pH 7.0) followed by dehydrating in a graded series of acetone (10, 30, 50, 70, 90, 100%) on ice for 15 min for each step. Samples were then subjected to critical-point drying with liquid CO<sub>2</sub> (CPD 30, Bal-Tec, Leica Microsystems, Wetzlar, Germany). Dried samples were covered with a gold film by sputter coating (SCD 500, Bal-Tec) and examined in a Zeiss Merlin field emission scanning electron microscope (FESEM) at an acceleration voltage of 15 kV. Imaging was performed with the high-efficiency SE2 Everhart-Thornley detector and Inlens-SE detector in a 50:50 ratio. Contrast and brightness of images was adjusted by applying Adobe Photoshop CS5.

### Computational sequence alignment

Amino acid sequences were aligned using MegAlign software tool provided by DNASTAR Lasergene 7.1 program according to the Jotun Hein Alignment with a gap penalty of 11 and a gap length penalty of 3.

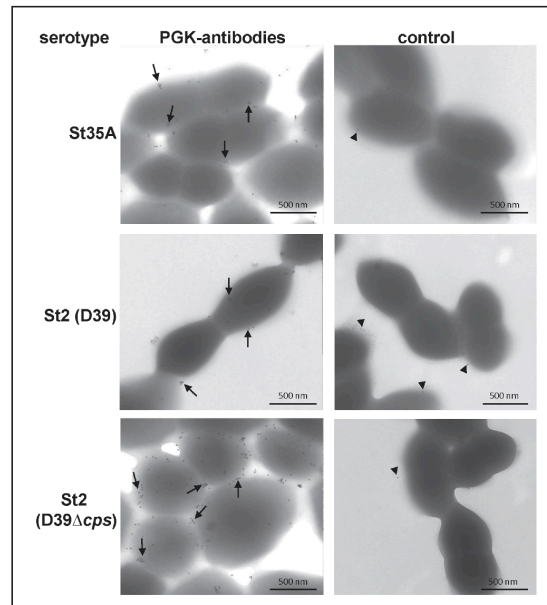
### Statistical evaluation

All experiments were performed in at least three independent assays, each in triplicate and the data were expressed as mean ± standard deviation. Statistical significances were analysed by the two-tailed unpaired Student's t-test; p-values of < 0.05 were considered as statistically significant.

## Results

### PGK is expressed on the bacterial surface

A prerequisite for environmental interactions is the exposition of pneumococcal binding proteins on the bacterial surface. After generation and purification of antibodies specifically recognising pneumococcal PGK, surface expression was visualised by transmission electron microscopy using immune gold labelling technique. PGK was detected in the polysaccharide matrix of the capsulated serotype 2 *pneumococcus* (D39) and also on the cell wall of the less encapsulated serotype 35A strain (► Figure 1). Interestingly, the amount of detected PGK was highest on the surface of the serotype 2-derived capsule-deletion mutant (D39Δcps) (► Figure 1), indicating a preferential location of PGK on pneumococcal cell wall.



**Figure 1: Transmission electron microscopic localisation of pneumococcal PGK (TEM).** PGK was detected on the surface of less capsulated serotype 35A, capsulated serotype 2 (D39) and the capsule-deficient serotype 2 derivative D39Δcps using PGK-specific antibodies and 15 nm protein A-gold particles. Black dots mark the detection of PGK. The controls represent incubation of the different pneumococcal serotypes with gold particles and indicate only minor unspecific background detection. Length of the size bars indicate 500 nm.

### Identification of PGK as plasminogen binding protein

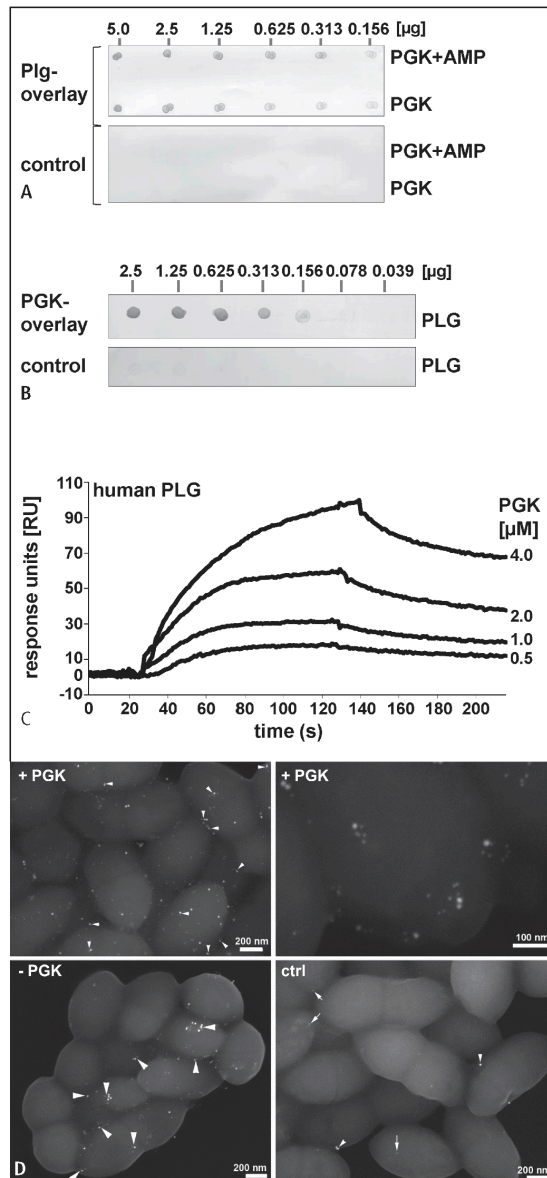
We have previously shown that *S. pneumoniae* binds human PLG (14). In order to analyse the role of PGK as PLG-binding protein, we produced pneumococcal PGK in *E. coli*. After cloning and expression of pneumococcal PGK in *E. coli*, dot spot overlay analyses identified a dose dependent PLG binding of purified PGK alone or substituted with AMP-PNP as stabilising agent (► Figure 2A). Interaction between PGK and PLG was also detected in a reverse PGK overlay after immobilisation of human PLG (► Figure 2B). In both dot blot overlay analyses, no unspecific binding of antibodies was detected (► Figure 2A, B control). The use of Surface Plasmon Resonance (SPR) confirmed binding of PLG to pneumococcal PGK, which was covalently coupled to CM5 sensor chips in concentrations of 4.0 μM, 2.0 μM, 1.0 μM and 0.5 μM. Evaluation of binding kinetics by determination of dissociation constants using 1:1 Langmuir binding algorithm indicated a nanomolar binding affinity for the interaction of PGK with human PLG ( $K_D = 1.3 \times 10^{-9}$  M, ► Figure 2C). The parameters describing binding dynamics are presented in ► Table 1. Scanning electron microscopy was performed after incubation of D39Δcps bacteria with PLG-gold nanoparticles (15 nm in size) in parallel to detection of sur-

face-exposed PGK with specific antibodies and gold-labelled protein A/G (10 nm in size). Electron microscopic visualisation detected clusters of both signals, indicating *in vivo* colocalisation of PGK and PLG (► Figure 2D, arrowheads). Preincubation of bacteria with PGK increased the amount of surface associated PGK-PLG complexes (► Figure 2D, +PGK in two magnifications). The control indicates only marginal unspecific background signals after incubation with PGK-preimmune serum and 15 nm gold

particles (control). The colocalisation studies confirmed localisation and reassociation of PGK on the bacterial surface and its function as PLG binding protein.

### PGK-mediated plasmin activity and fibrin degradation

Plasmin activity analyses were conducted in order to determine conversion of PGK-bound PLG to plasmin. Plasmin-mediated cleavage of the chromophore S-2251 was monitored in time intervals of 5 min (► Figure 3A). After incubation of PGK-coated Dynabeads® with PLG, supplementation of either tPA (line with rhombus) or uPA (line with x) resulted in an exponentially enhanced plasmin activation reaching a plateau after 25 min with tPA ( $1.27 \pm 0.12$ ) and after 40 min with uPA ( $1.2 \pm 0.21$ ) (► Figure 3A). In contrast, only minor activity was determined in the absence of an activator (PGK-PLG, line with dots). A weak linear increasing plasmin activity was observed after PLG incubation of BSA-coated Dynabeads® in the presence of either tPA (dotted line with triangles) or uPA (broken line with squares) indicating minor unspecific PLG-binding to BSA-coated beads. Both, PGK-coated Dynabeads® and BSA-coated Dynabeads® showed no residual plasmin activity without PLG-incubation in presence of either tPA or



**Figure 2: Biochemical characterisation and electron microscopic visualisation of PGK-plasminogen interaction.** A) Blot overlay with human PLG after immobilisation of recombinant PGK on nitrocellulose at indicated concentrations in absence or presence of the ATP-analogue AMP-PNP as stabilising substrate. PLG binding was detected with PLG-specific antibodies and peroxidase-conjugated secondary antibodies. Without PLG incubation, no unspecific binding of antibodies was detected (control). B) Blot overlay with recombinant PGK after immobilisation of human PLG. Binding was detected with PGK-specific antibodies and peroxidase-conjugated secondary antibodies. Immunoblot with PGK-specific antibodies indicated no unspecific antibody recognition (control). C) Surface plasmon resonance (SPR) measurements of PGK interaction with human PLG. PLG was immobilised on a BIAcore™ CM5 sensor chip and recombinant PGK was used as analyte. Changes in Plasmon resonance are shown as relative response units (RU). PGK was injected in concentrations of 4.0  $\mu$ M, 2.0  $\mu$ M, 1.0  $\mu$ M and 0.5  $\mu$ M. Sensograms demonstrate concentration dependent binding of PGK to human PLG and were used for evaluation of binding kinetics. The blank run was subtracted from each sensogram. BIAcore results representing the association rate constant ( $k_a$ ), the dissociation rate constant ( $k_d$ ), the affinity constant ( $K_D$ ) and the  $\chi^2$  values are shown in Table 1. D) Field scanning electron microscopy (FESEM) of PGK-PLG colocalisation. D39 $\Delta$ cps bacteria were pre-incubated with recombinant PGK protein and incubated with PLG-gold nanoparticles (15 nm). Surface associated PGK was detected with specific antibodies and protein A/G gold nanoparticles (10 nm). Visualisation by FESEM on the two upper pictures (+ PGK in two different magnifications) detected both nanoparticles in clusters on the bacterial surface, indicating colocalisation of PGK and PLG (arrow heads). Similar colocalisation clusters but less amount of PGK-detecting nanoparticles were visualised on bacterial surface without PGK preincubation (-PGK). Only marginal signals were detected after incubation of bacteria with preimmune serum and 15 nm gold nanoparticles in a control experiment (ctrl). Arrows mark unspecific background binding of PGK nanoparticles (10 nm) and arrow heads point to gold particles (15 nm).

uPA (► Figure 3A). In contrast, PLG-coated beads exhibited similar plasmin activity in presence of either tPA or uPA as PGK-coated beads after PLG in presence of tPA or uPA (resulting in absorbance at 405 nm of  $1.77 \pm 0.05$  (with tPA) and  $1.695252 \pm 0.03$  (uPA, data not shown)).

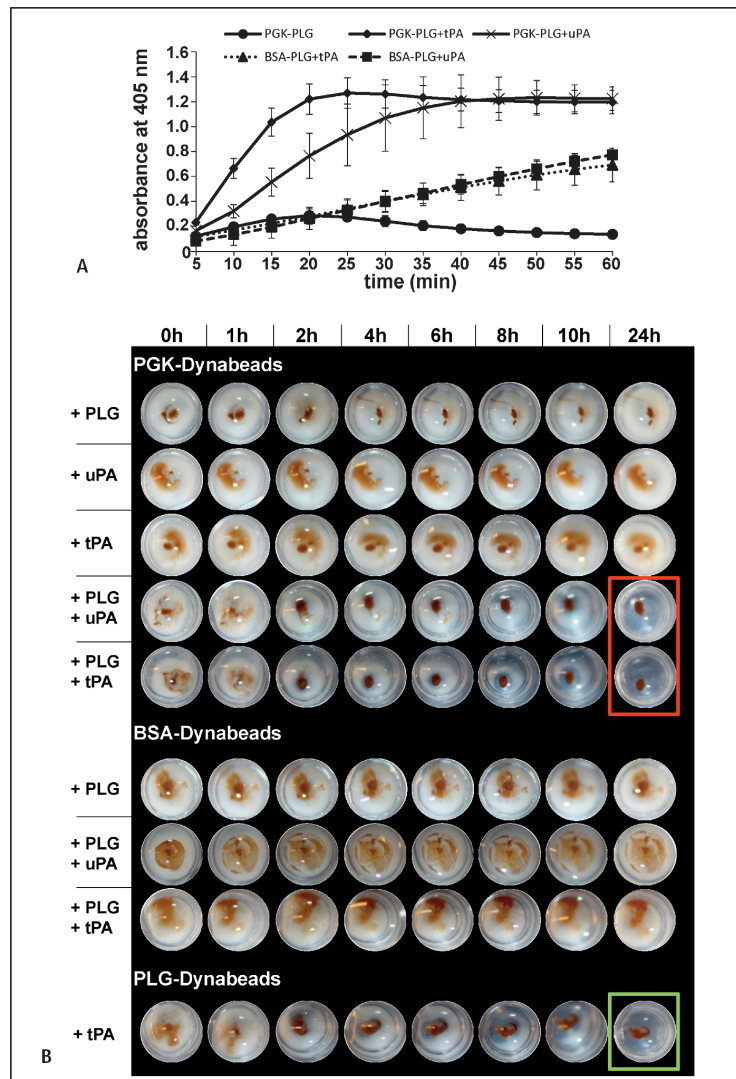
Incubation of synthetically generated fibrin aggregates with PGK-coated Dynabeads® after PLG incubation in presence of tPA or uPA resulted in a stepwise dissolution of the solid, white fibrin matrix into clear liquid. Photographic illustration of fibrin aggregates generated on 24-well transwell insets indicated the start of visible fibrin dissolution 4 h after incubation with tPA or uPA as activator (► Figure 3B, PGK-Dynabeads® + PLG + tPA and PGK-

Dynabeads® + PLG + uPA). After 6–8 h the process of fibrin degradation has reached the bottom of the transwell inset. The complete dissolution of the turbid fibrin matrix into a clear solution was monitored after 24 h of incubation (► Figure 3B, PGK-Dynabeads®, red frame). Fibrin dissolution was also observed with PLG-coated Dynabeads® in the presence of tPA as positive control (► Figure 3B, PLG-Dynabeads®, green frame). In contrast no visible fibrin dissolution was monitored after incubation of the fibrin meshwork with PGK-coated Dynabeads® alone or without an activator and after incubation with BSA-coated Dynabeads®.

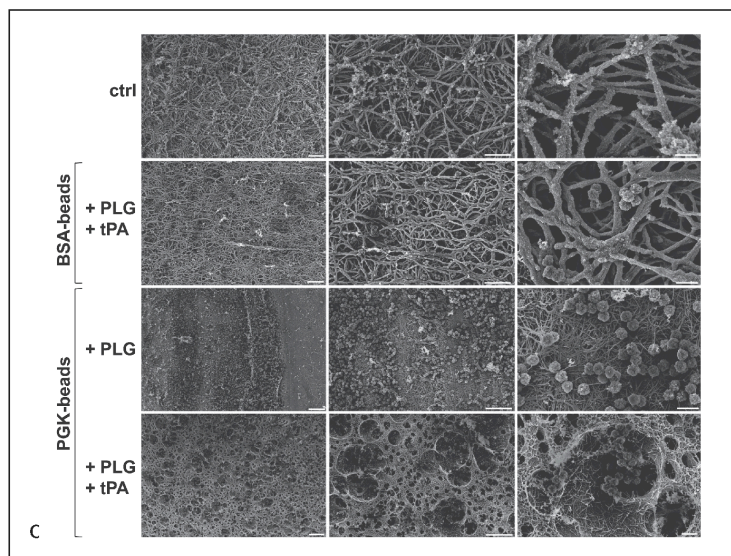
Electron microscopic visualisation of fibrin degradation at three different magnifications by PLG-preincubated PGK-Dyna-

**Figure 3: Plasmin activation and fibrin degradation by PGK-coated Dynabeads®.**

A) Generation of plasmin was determined by ELISA-measurement with the chromogenic substrate S-2251 in 5 min intervals up to 60 min. PGK-coated Dynabeads® were incubated with PLG alone (PGK-PLG, line with dots) and with tPA (PGK-PLG + tPA, line with rhombus) and with uPA (PGK-PLG + uPA, line with x). Background activity of unspecific PLG binding to BSA-coated Dynabeads® was determined after incubation with tPA (BSA-PLG + tPA, dotted line with triangles) or uPA (BSA-PLG + uPA, broken line with squares). Results are presented as mean values of triplicates with standard deviation. B) Aggregated fibrin thrombi were generated by incubation of fibrinogen with thrombin over night resulting in milky white solid matrix. The matrix was incubated with PGK-coated Dynabeads® or BSA-coated beads after incubation with PLG alone or with PLG and tPA or with PLG and uPA. The magnetic beads are visible as little brown spots on the white fibrin matrix. Dissolution of fibrin matrix into clear liquid was photographically visualised in a time series for up to 24 h. Red frame indicates total fibrin dissolution by PGK-coated Dynabeads® after PLG incubation in presence of tPA or uPA. Incubation of fibrin matrix with PLG-coated beads in presence of tPA was applied as positive control and the green frame marks total fibrin dissolution. Figure shows representative examples.







**Figure 3:** C) Electron microscopic visualisation of fibrin thrombi generated on cover slips after incubation for 3 h with PLG-preincubated PGK-coated Dynabeads® and tPA (PGK-beads, +PLG, +tPA). Three different magnifications for each incubation set up are illustrated. Size bars of pictures at left hand site represent 20 μm, of pictures in the mid position 10 μm and on the right hand site 2.0 μm. The first row depicts the intact aggregated fibrin matrix after 3 h of incubation in PBS as control (ctrl). No fibrin dissolution is visible after incubation of the fibrin thrombi with PLG-preincubated BSA-coated Dynabeads® in presence of tPA (BSA-beads, +PLG, +tPA) or with PLG-preincubated PGK-beads without activator (PGK-beads, +PLG).

beads\* indicated a substantial dissolution of the regular fibrin network (► Figure 3C). The generated plasmin activity dissolved the strong and associated fibrin bundles (shown in ► Figure 3C as control) leaving a crater-like hole around the bead particles containing cleaved and shortened fibrin stubs (► Figure 3C, PGK-beads, +PLG, +tPA). No degradation activity was observed after incubation of the fibrin thrombi with PLG-preincubated PGK-beads without activator (PGK-beads, +PLG) or with PLG-preincubated BSA-beads in presence of tPA (BSA-beads, +PLG, +tPA). The electron microscopic pictures confirm the generation of plasmin activity after PLG binding to surface-displayed PGK in presence of tPA and demonstrate the promoted degradation of aggregated fibrin thrombi.

#### Plasminogen binding is species-unspecific

The pathogenic potential of *S. pneumoniae* is not restricted to humans, but also to other mammals such as horses, dolphins, and guinea pigs (31). Results of dot spot analyses using human and murine PLG as ligand (► Figure 4A) or analyte (► Figure 4B), respectively, demonstrated a similar binding strength of pneumococcal PGK to both, human and murine PLG. Moreover, after immobilisation of murine PLG onto CM5-Sensorchips, SPR analysis revealed a concentration-dependent binding of pneumococcal PGK (► Figure 4C). The respective dissociation constant of  $K_D = 4.69 \times 10^{-7}$  M was lower compared to the dissociation constant obtained for the interaction between PGK and human PLG ( $K_D = 1.3 \times 10^{-9}$  M) (► Figure 4C, D and ► Table 1). Comparison of single kinetic parameter indicated similar association rates of  $8.39 \times 10^3$  M<sup>-1</sup> for human PLG and  $7.98 \times 10^3$  M<sup>-1</sup> for murine PLG, but a higher dissociation rate for the interaction of PGK with murine PLG ( $1.06 \times 10^{-5}$  s<sup>-1</sup> for human PLG and  $3.75 \times 10^{-3}$  s<sup>-1</sup> for murine

PLG, ► Table 1). These results show that interaction of pneumococcal PGK with murine PLG occurs specific but with lower affinity and indicate a species-independent interaction.

#### Identification of angiostatin as PGK-ligand

Cleavage of PLG with elastase results in generation of different PLG fragments including the N-terminal kringle domains 1–3 and 1–4 representing angiostatin and the C-terminal mini-PLG containing kringle domain 5 and the enzymatic domain (8) (► Figure 5A). Dot spot overlay analyses demonstrated a dose-dependent binding of full-length PLG, angiostatin (ANG), and kringle domains 1–3 (Kringle 1–3) to pneumococcal PGK (► Figure 5B). Furthermore, immunoblots confirmed that PLG-specific polyclonal antibodies recognised all immobilised PLG fragments (► Figure 5B, control 1) but no unspecific binding to PGK was detected (► Figure 5B, control 2). In contrast, no PGK binding was detected using mini-PLG as analyte (► Figure 5A, B). Therefore the binding site for pneumococcal PGK in PLG could be narrowed down to the first three kringle domains of PLG representing angiostatin.

#### Localisation of plasminogen binding sites in PGK

In order to identify the PLG binding sites within the PGK protein, a spot membrane-based peptide array representing the whole amino acid sequence of pneumococcal PGK was conducted. The PGK peptides were 16 amino acids in length with an offset of four amino acids. Overlay analyses with human PLG indicated the presence of two quaternary groups of peptide spots (BS1: GKKVLVRVDFNVPLKD, BS2: RAILFSHLGRVKEESD) with increasing PLG binding intensity (► Figure 6A). Sequence mapping

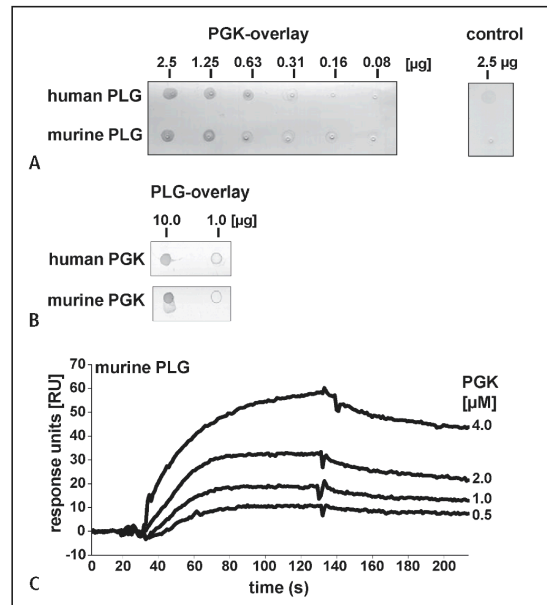
indicated that the amino acid motifs "GKKV" (BS1) and "RAIL" (BS2) are crucial for PLG binding. The importance of the single amino acids in PLG binding was investigated in a second peptide array comprising the 16mers of BS1 and BS2 with a systematical alanine exchange at each subsequent amino acid position. After PLG overlay, a significant reduction of binding signal was detected for two peptide spots with substitutions of the lysines in BS1 (peptide no. 3 and 4) resulting in the peptide sequence "GAKVLVRVDFNVPLKD" (no. 3) and "GKAV-LVRVDFNVPLKD" (no. 4, ►Figure 6B, C). A further loss of binding signal was identified in BS2 after alanine substitution of an arginine (peptide no. 19) resulting in the peptide sequence "AAILFSLGRVKEESD" (no. 19, ►Figure 6B, C). The peptide spot array analyses detected two putative PLG binding sites in the amino-terminal domain of PGK and identified two lysines of BS1 and an arginine in BS2 as crucial for interaction.

### Structural determinants of PLG binding by PGK

Three-dimensional structure of pneumococcal PGK was solved by X-ray crystallography at 1.78 Å resolution (►Table 2). PGK folds into two globular domains, the N-domain and C-domain, consisting of residues 1–191 and 192–398, respectively (►Figure 7A). The 10 C-terminal residues cross back to the N-terminal domain, via a short helical segment leaving the Cα-atoms of the N- and C-terminal residues only 8 Å apart. Each domain folds as a  $\alpha/\beta$  domain consisting of five parallel  $\beta$ -strands (in N-domain), and four parallel  $\beta$ -strands (in C-domain); both sheets are sandwiched between four long  $\alpha$ -helices, two to a side. In addition, C-domain has two antiparallel  $\beta$ -strands above the C-termini of the parallel  $\beta$ -strands. A search for structurally PGK related proteins performed by PDBe server identified the PGKs from *Geobacillus stearothermophilus* (PDB code 1PHP, Z score of 20.0 and rmsd of 1.33 Å for 387 Cα-atoms) and from *Bacillus anthracis* (PDB code 3B2B, Z score of 19.7 and rmsd of 1.31 Å for 382 Cα-atoms) as the most closely related known three-dimensional (3D) structures.

The binding site for the ATP-analog AMP-PNP is located in the C-domain. The nucleotide is oriented with the adenine ring distal to the N-domain and the triphosphate moiety adjacent to the interdomain cleft and stabilised by many polar interactions, some of them mediated by water molecules (Suppl. Figure S1, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)). The 3-phosphoglycerate binding site is located at the active site of the N-terminal domain. This site was not occupied in our pneumococcal PGK structure.

Two PLG binding sites, BS1 (presenting the  $^{13}\text{GKKV}^{16}$  motif) and BS2 (presenting the  $^{53}\text{RAIL}^{56}$  motif), were identified by spot membrane-based peptide array. Crystal structure of pneumococcal PGK reveals that both sites are indeed placed in the same locus defining a single PLG binding site. This region is located in the N-terminal domain at the opposite side of the catalytic site and presents a large number of well-paired surface charges (►Figure 7B). Besides the Lys14, Lys15, and Arg53 from BS1 and BS2; PLG binding site presents other positively charged residues (Lys12) and negatively charged residues (Asp10, Glu101,



**Figure 4: Dot spot analyses and surface plasmon resonance measurements of PGK-plasminogen interactions.** A) PGK-overlay of a dot spot membrane containing PLG-protein spots derived from different species. Human and murine PLG were immobilised on a nitrocellulose membrane in serial 1:2 dilution starting from 2.5 µg down to 0.08 µg. After blocking of the membrane with 10% skimmed milk, protein overlay was performed with 40 µg of purified PGK. Immunoblot with PGK-specific antibodies indicated only marginal unspecific antibody binding to 2.5 µg immobilised human or murine PLG (control). B) PLG-overlay of a dot spot membrane with PGK protein spots. In a similar approach human and murine PLG were used as analytes in solution and bound to immobilised PGK protein (10.0 µg and 1.0 µg). C) SPR-analysis of PGK interaction with murine PLG. Murine PLG was immobilised on a BIAcore™ CM5 sensor chip and purified recombinant PGK was used as analyte. Changes in plasmon resonance are shown as relative response units (RU). PGK was injected in concentrations of 4.0 µM, 2.0 µM, 1.0 µM and 0.5 µM. Corresponding sensograms were used for evaluation of dissociation constant according to 1:1 Langmuir binding model. BIAcore results representing the association rate constant ( $k_a$ ), the dissociation rate constant ( $k_d$ ), the affinity constant ( $K_D$ ) and the  $\chi^2$  values are shown in Table 1.

**Table 1: BIAcore parameter of binding dynamics.**

Ligand	Analyte	$k_a$ (1/Ms)	$k_d$ (1/s)	$K_D$ (M)	$\chi^2$
Human PLG	PGK	$8.39\text{e}^3$	$1.06\text{e}^{-5}$	$1.27\text{e}^{-9}$	16.2
Murine PLG	PGK	$7.98\text{e}^3$	$3.75\text{e}^{-3}$	$4.69\text{e}^{-7}$	4.4

Glu108, Asp109, Asp144, and Glu170) (►Figure 7B). This paired charge distribution has been also observed for other PLG binding proteins such as enolase (32). Sequence analysis among PGK from different species suggests that the regions that we im-

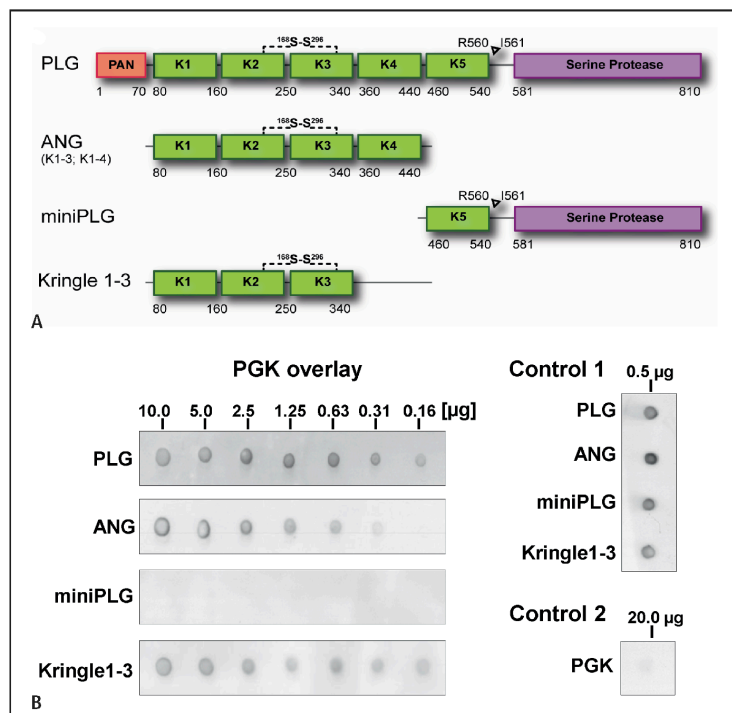
plicate in the PLG binding are not conserved in bacterial PGKs other than streptococci (Suppl. Figure S2, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)).

A computational model of the PGK:Kringles complex has been generated based on our PGK crystal structure and the reported three-dimensional structures of kringles (see *Methods*). This model (► Figure 7C) faces the PGK PLG binding site, with the lysine-binding site of K2 and reveals a perfect charge complementarity between the pneumococcal protein and human kringle structure. Interestingly, PGK:Kringles complex reveals that Lys14 residue from PGK would play a central role by interacting with both Trp residues of the Lysine-binding site through a cation- $\pi$  interaction and making a salt bridge interaction with Glu59 from kringle (► Figure 7C). Around this central interaction many polar and electrostatic interactions are predicted between PGK and kringle structures. Very recently the crystal structure of inactive conformation full-length human PLG comprising the Pan-apple (PAP) domain, five kringle domains (K1–5), and a serine protease (SP) domain, has been reported (33). Interestingly, the interaction of K2 with the SP domain observed in human PLG is extremely close to that predicted for the PGK:K2 complex (Suppl. Figure 3, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)). In both cases a Lys residue is inserted in the Trp hydrophobic cavity of kringle, while establish a salt bridge interaction with acidic residues from the kringle. The central interaction mediated by the Lys residue is complemented by many salt bridge interactions between kringle

and SP/PGK charged residues (Suppl. Figure S3, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)).

### Binding of PGK to tissue-type plasminogen activator

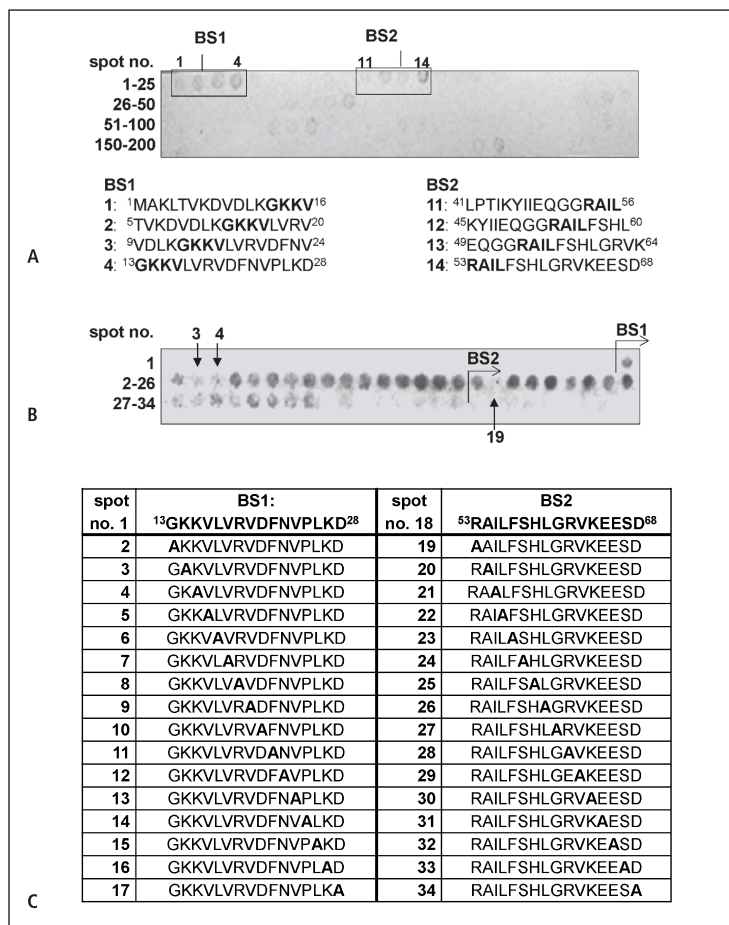
Besides PLG, other human-derived proteins involved in fibrinolysis like PLG activators comprise kringle domains. Two PLG activators serve as early mediators of fibrinolysis and degrade fibrin. Tissue-type plasminogen activator (tPA) possesses two such structural units (34) and urokinase plasminogen activator (uPA) presents one kringle module (35) (Suppl. Figure S5A, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)). Binding of tPA has already been reported to some non-streptococcal species, whereas no streptococcal binding to tPA and no bacterial binding to uPA has been demonstrated so far (36). The 3D structures of kringle 2 domain of tPA (34) and angiotensin (kringle domains 1 to 3) (37) have been reported. Structural superimposition of these structures on modelled PGK:K2 complex (► Figure 8A) and molecular dynamic simulations (Suppl. Figure S6, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)) reveal that despite structural differences between these modular proteins, all of them could be fitted on the PLG binding site found in PGK. To rule out this hypothesis dot blot binding analyses with immobilised human PLG, tPA and uPA were performed. Our results confirmed a dose-dependent binding of pneumococcal PGK to PLG and tPA but not to uPA (► Figure 8B).



**Figure 5: Dot spot analyses of PGK interaction with human plasminogen different plasminogen derivatives.** A) Schematic model of PLG domain structure and used PLG derivatives. The PAN-apple domain is shown in red, the kringle domains are depicted in green, and the serine protease domain is shown in violet (K, Kringle; PAN, PAN-apple (PAP)-domain; ANG, angiotensin, miniPLG, mini-plasminogen). B) Dot Spot analyses of PGK binding to different PLG derivatives. Pneumococcal PGK was immobilised in different amounts ranging from 10.0 μg down to 0.16 μg onto a nitrocellulose membrane and human plasminogen (PLG) and the plasminogen derivatives angiotensin (ANG), representing kringle domains 1–3 and 1–4 (K1–3; K1–4), mini-plasminogen (miniPLG), representing the last kringle domain (K5), and the enzyme domain, and kringle domains 1–3 (Kringle 1–3) were used as analytes in overlay studies. Detection of binding signal was performed with PLG-specific antibodies. PGK bind to each of the PLG proteins and derivatives with exception of mini-plasminogen. Control 1 represents equal detection intensity of the different PLG derivatives by PLG-specific antibodies. Control 2 shows no unspecific background binding of PLG antibodies to PGK.

**Figure 6: Determination of plasminogen binding sites in PGK by peptide spot array.**

A) PLG-overlay of PGK-peptide spot array. A peptide spot membrane of the 398 amino acid sequence of PGK divided into overlapping peptides of 16 amino acids in length with an offset of four amino acids was analysed for PLG binding. Twenty-five peptides were immobilised in one line. Binding of PLG was detected in spots 1 to 4 (BS1, first box) representing the peptide sequence from 1 to 28 and in spots 11 to 14 (BS2, second box) representing the peptide sequence from 41 to 68. The crucial four amino acids were marked in bold. Reactivity of other spots was due to unspecific binding of plasminogen-antibodies and secondary antibodies. B) Alanine-amino acid exchange analyses of PLG-binding sites BS1 and BS2 of PGK. PLG overlay analyses of a peptide spot membrane representing a systematic alanine amino acid exchange of the 16mer peptide sequences BS1 and BS2 at each single amino acid position. Alanine exchange at position 3 and 4 of BS1 and at position 19 representing amino acid 2 of BS2 leads to loss of PLG-binding activity C) Amino acid sequences of the immobilised peptide spots after alanine exchange. Two putative PLG binding sequences were identified BS1: GKKVLVRVDFNVPLKD, BS2: RAILFSLGRVKEESD with the lysines of BS1 (in bold) and the arginine in BS2 (in bold) as crucial amino acids.



In addition to structural analysis, a sequence alignment between tPA and uPA was carried out to explain the observed differences in binding activity for tPA and uPA (Suppl. Figure S5B, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)). Interestingly, one of the two conserved Trp residues forming the binding site in PLG kringle domains and in tPA (Trp235 in PLG and Trp72 in tPA) is substituted by Val123 in uPA. Of special relevance is a substitution of Asp57 (tPA) by Arg108 (uPA) (Suppl. Figure S5C, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)) as this Asp residue is involved in the critical salt-bridge interaction with Lys residue of PGK. This interaction, observed in the crystal structure of inactive PLG and also in our modelled complexes between K2:PGK and tPA:PGK, would not be possible between PGK and uPA. In summary, differences in uPA binding-site conformation together with the presence of Arg108 residue would impair the interaction with PGK.

Plasmin activity analysis was performed with PGK-coated Dynabeads®, preincubated with tPA. Subsequent incubation with PLG resulted in a fast and substantial plasmin generation (PGK-tPA + PLG, line with triangles, ► Figure 8C) reaching an absorbance of  $1.06 \pm 0.02$  after 60 min of incubation. In contrast, only marginal plasmin activity was monitored after incubation of PGK-coated beads with uPA prior to PLG supplementation (PGK-uPA + PLG, line with x, ► Figure 8C). Only minor conversion of PLG into plasmin was determined using PGK-beads after incubation with tPA (PGK-tPA, broken line with dots) or uPA (PGK-uPA, broken line with rhombus) alone. Plasmin activity was absent after incubation of tPA- or uPA- preincubated BSA-beads, indicating no unspecific binding of plasminogen activators to the BSA-coated beads (data not shown). These results confirm the direct interaction between PGK and tPA and demonstrate the effective conversion of PLG into plasmin by tPA bound to PGK protein.



Table 2: Data collection and refinement statistics.

	PGK
<b>Data collection</b>	
Space group	P2 <sub>1</sub>
a, b, c (Å)	40.35, 78.23, 59.03
$\beta$ (°)	96.34
T (K)	100
X-ray source	Synchrotron (ID14-2)
Wavelength (Å)	0.933
Resolution (Å)	46.94 (1.95–1.78)
Total no. reflections	129733 (18010)
No. unique reflections	34968 (5054)
Redundancy	3.7 (3.6)
Completeness (%)	99.9 (100)
Average I/ $\sigma$ (I)	27.5 (2.2)
R <sub>merge</sub>	0.08 (0.59)
<b>Refinement statistics</b>	
Resolution (Å)	46.93 (1.95–1.78)
R <sub>work</sub> /R <sub>free</sub>	0.14/0.19
Non-hydrogen atoms	3420
Protein	2957
ANP-PNP	31
Solvent	404
Glycerol	24
Magnesium	4
<b>B-factor values (Å)<sup>2</sup></b>	
Protein	16.06
ANP-PNP	15.03
Solvent	26.37
Glycerol	30.01
Magnesium	24.14
Rmsd bond length (Å)	0.01
Rmsd bond angles (°)	1.35
PDB code	3ZLB
Values in parenthesis correspond to the highest resolution shell. R <sub>merge</sub> = $\sum (I - \langle I \rangle) / \sum I$ .	

## Discussion

Interaction of *S. pneumoniae* with host derived PLG has been characterised as powerful virulence mechanism facilitating dissemination within the host tissue and transmigration of tissue barriers by subversion of the proteolytic potential (5, 6). Our work reveals that the pneumococcal glycolytic enzyme PGK has to be classified as an additional “moonlighting protein” that is expressed on the

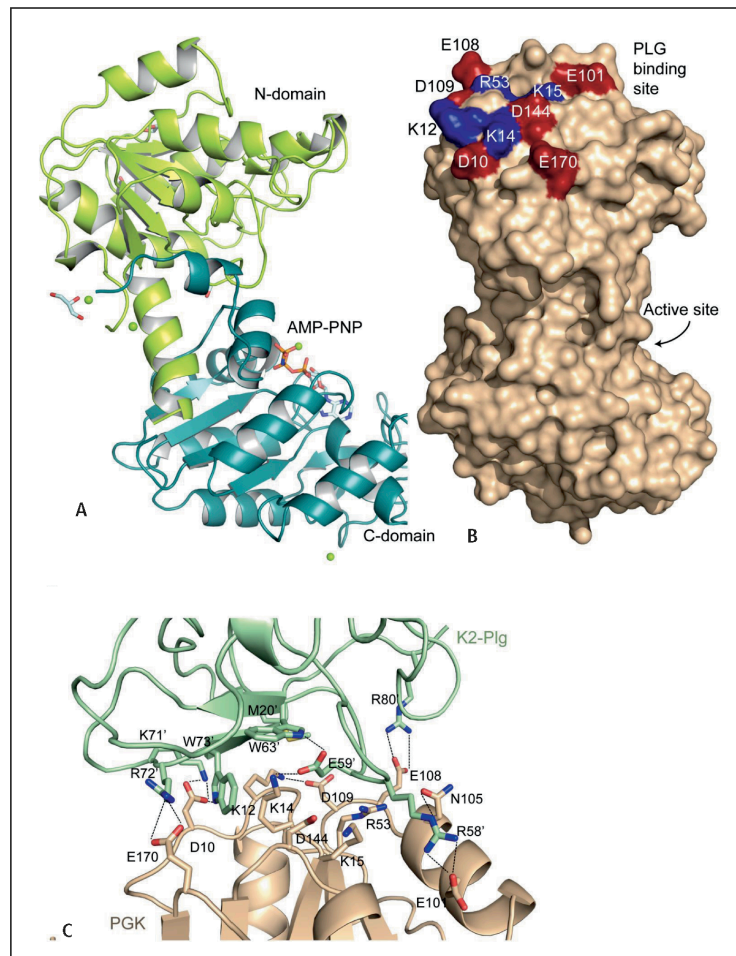
surface and interacts with host proteins of the fibrinolytic system. PLG binding by bacterial PGK has also been reported for oral streptococci (*S. anginosus* and *S. oralis*) (38), and *S. agalactiae* (39, 40). In accordance with our former studies on enolase, electron microscopic analysis demonstrates surface localisation and surface reassociation of PGK. Furthermore, FESEM visualised colocalisation of PGK with PLG, confirming PGK as PLG binding protein. SPR-binding analyses of the interaction between the pneumococcal enolase and PLG suggested a conformational change resulting in a two-step dissociation (20), while SPR evaluation of PGK-PLG interaction followed a Langmuir 1:1 binding model. The calculated dissociation constant of PGK-PLG binding of  $K_D = 1.3 \times 10^{-9}$  M was in a similar high-affinity range compared to the dissociation constants reported for the interaction of the pneumococcal enolase with PLG ( $K_{D1} = 0.55 \times 10^{-9}$  M and  $K_{D2} = 86.3 \times 10^{-9}$  M) (20) and for *S. pyogenes* enolase SEN with PLG (41, 42). Determination of a dissociation constant of  $4.69 \times 10^{-7}$  M for the binding of murine PLG to PGK confirmed the results of dot blot binding studies and indicated a species-unspecific interaction, but with lower affinity. Although computational analyses of amino acid sequence homology revealed similarity of 78.8% between human PLG (Ac.no. : AAA36451.1) and murine PLG (Ac.no. AAA50168.1), the reduced binding affinity to murine PLG could be the result of residual differences in the amino acid sequence.

Structural determination of PGK indicated that PLG-binding site is located at the N-terminal part of PGK in opposite location of the catalytic site, thereby most probably unaffected conformational changes or enzymatic activities. This is also confirmed by the Langmuir 1:1 binding model applied for SPR analyses, representing PGK-PLG interaction without conformational changes.

In contrast to several PLG binding proteins i.e. eukaryotic enolase that interact with PLG lysine binding sites via C-terminal lysine residues, internal lysine-containing binding motifs have been characterised as major binding sites for several bacterial PLG-binding proteins including pneumococcal enolase (5) and the M-protein of group A streptococci (PAM) (43). Alignment of peptide sequences of the pneumococcal PGK with enzymes of other oral streptococci like *S. oralis* and *S. anginosus* and with *S. agalactiae* revealed up to 98% sequence identity (Suppl. Figure S4, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)), implying similar biochemical binding mechanisms. Recently, lysine-rich PLG binding sites were identified via peptide mapping in the PGK protein of *S. agalactiae* (40). Interestingly, the pneumococcal PGK peptide sequences which constitute highest homology to the identified PLG-binding sites of *S. agalactiae* PGK do not react with PLG in a similar peptide spot array. Nevertheless, a structural overlay of the PGKs of *S. pneumoniae* and *S. agalactiae* indicated a localisation of the binding site (<sup>126</sup>KKESKNDEE<sup>134</sup>) from GBS (40) within the PLG binding region formed by the pneumococcal PLG-binding sites BS1 and BS2. Therefore, the results of peptide spot arrays and structural mapping point to a similar binding chemistry of pneumococcal and group B streptococcal PGK and illustrate the crucial contribution of the identified BS1 and BS2. Lastly, Boone and Tyrrell suggested a glutamic residue as crucial for PLG binding of *S. agalactiae* PGK (40). However, this position is occupied by a pro-



**Figure 7: Crystal structure of pneumococcal PGK and modelled interaction with plasminogen.** A) Crystal structure of the PGK protein. The N-domain (residues 1–191) is depicted in green, C-domain (192–398) in blue, glycerol molecules and AMP-PNP are depicted in sticks.  $Mg^{2+}$  ions are drawn as green spheres. B) Location of the PLG binding site on PGK molecular surface. PGK (brown) is oriented as in panel A. Paired charged distribution is found in the PLG binding site. Negatively charged residues (Asp10, Glu101, Asp109, Asp144 and Glu170) are depicted in red and positively charged residues (Lys12, Lys14, Lys15 and Arg53) in blue. C) Computational model of PGK in complex with plasminogen K2 domain. Residues involved in the predicted interactions between PGK (wheat colour) and K2 (green) are drawn as sticks. Polar interactions are represented as dotted lines.

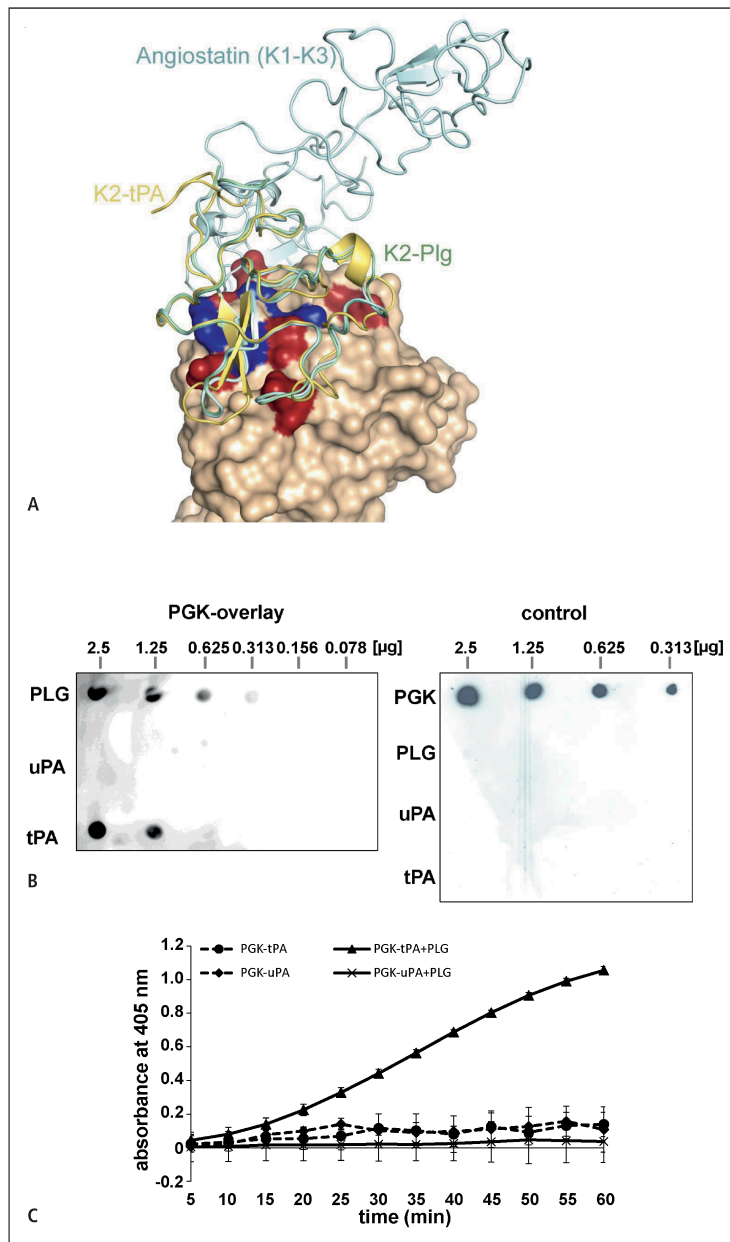


line in both the *S. pneumoniae* PGK and in the *S. anginosus* PGK and by a serine in *S. oralis* (Suppl. Figure S4 marked in blue, available online at [www.thrombosis-online.com](http://www.thrombosis-online.com)), all of them reported to bind PLG (40). These data indicate that the amino acid at position 133 is not likely critical for PLG binding.

PLG is known to interact with its ligands via pre-formed lysine-binding sites within the N-terminally located triple-disulfide-bonded kringle domains (7). Reported structures of different kringle domains (44–48) delineate the basic characteristics of the binding pocket (named lysine-binding site) of kringles (49, 50), which consists of a groove at the surface of the molecule, formed by two tryptophan side chains, flanked by charged residues. In contrast to the interaction of the *S. canis* M protein SCM with mini-PLG (51, 52), dot spot results demonstrate binding of the pneumococcal PGK to angiostatin, which represents the kringle domains 1–3. Likewise, interaction of two internal peptide sequences (a1 and a2) of the *S. pyogenes* M protein PAM to kringle

domain 2 of plasminogen has been reported (53). Crystal structure of pneumococcal PGK shows the presence of a patch of exposed lysine-rich residues that might mediate interaction with the PLG K2 as observed for the interaction of K2 with the SP domain in the crystal structure of inactive PLG (35).

Our structural model for the interaction between PGK and K2 (► Figure 7C and ► Figure 8A) pointed to a potential interaction with other kringle-containing proteins such as PLG activators tPA and uPA. Overlay analyses confirmed binding of tPA to pneumococcal PGK but not for uPA. This differential binding has been also observed in non-streptococcal bacteria (*E. coli* S-fimbria, *Haemophilus* and *Neisseria*) which bind tPA but not uPA (54). This is, to our knowledge, the first report of a tPA binding to *S. pneumoniae*. It is worth noting that up to now none of the tested streptococci (*S. pyogenes*, *S. equisimilis* and group G streptococci) showed binding for the PLG activators (55), indicating a unique binding of *S. pneumoniae* to tPA. The interaction with the amino-



**Figure 8: Binding of PGK to K2, angiostatin, tissue-type plasminogen activator (tPA) and plasmin activation.** A) Structural superimposition of K2, angiostatin and tPA on PGK-PLG binding site. The structures of kringle domain K2 (green), angiostatin (K1–3, light blue) and tPA (yellow) were fitted to the PLG binding site of the PGK N-terminus. B) Dot spot overlay analyses of PGK binding to PLG and PLG activators. Human PLG, uPA and tPA were immobilised on nitrocellulose at indicated amounts and PGK was employed for Dot Spot overlay analyses. Interaction of PLG and tPA with PGK was detected with PGK-specific antibodies and peroxidase-conjugated secondary antibodies, whereas no interaction with immobilised uPA was detected. Immunoblot with PGK-specific antibodies detected PGK but indicated no unspecific antibody binding to PLG, uPA and tPA (control). C) PGK-coated Dynabeads® were incubated with tPA or uPA alone (PGK-tPA, broken line with dots; PGK-uPA, broken line with rhombus) or with PLG (PGK-tPA + PLG, line with triangles; PGK-uPA + PLG, line with x). Generation of plasmin was determined by ELISA-measurement at 405 nm with the chromogenic substrate S-2251 in 5 min intervals up to 60 min. Results are presented as mean values of triplicates with standard deviation.

terminal kringle domains may allow a conformational opening and enhanced activation of the proteolytic active center of PLG. It has been demonstrated that recruitment of plasmin activity to the pneumococcal surface facilitates degradation of fibrin, thrombin and complex extracellular matrix, thereby providing enhanced dissemination through tissue barriers and entry into the vascular sys-

tem (5). In contrast to other streptococcal species, pneumococcal plasminogen binding proteins are not acting as direct plasminogen activators. But consistent with all pneumococcal PLG binding proteins identified so far, results of plasmin activity analysis indicated that PLG binding to PGK promotes conversion of PLG into plasmin via host-derived tPA or uPA. A remarkable and unique prop-

erty of PGK is the recruitment of tPA, which also leads to enhanced activation of plasminogen and subsequently to an effective degradation of fibrin thrombi. The synergistic immobilisation of both PLG and the host-derived activator tPA to the bacterial surface might promote the efficiency and speed of enzymatic PLG conversion via tPA at the local site of pneumococcal infection and may translocate fibrinolytic cascade directly to the bacterial surface in order to increase the resulting degradation activity, which may facilitate bacterial spread within host.

Considering that fibrinolysis also contributes to complex inflammatory processes, our results describe a new scenario in which this important human pathogen can subvert the host fibrinolytic machinery during the initial phase of bacterial colonisation enhancing both dissemination through tissue barriers and entry into the vasculature.

#### Acknowledgements

We thank Astrid Droege for technical assistance and Ina Schleicher (HZI, Braunschweig) for the preparation of probes for electron microscopy. We are also grateful to Uwe Schubert (Justus-Liebig University, Giessen) for purification of PLG and various derivatives and thank Andy Polok, Andreas Raschka, and Sandra Koch for technical assistance. The authors thank the ID14-2 beamline at the ESRF facility of Grenoble for access to synchrotron radiation.

#### What is known about this topic?

- *Streptococcus pneumoniae* binds human plasminogen via several surface-exposed binding proteins, but does not express an endogenous plasminogen activator.
- Recruitment of plasminogen to pneumococcal surface promotes conversion of plasminogen into proteolytic active plasmin.
- Bacterial-surface-bound plasmin activity leads to dissolution of fibrin thrombi and to degradation of extracellular matrix proteins of the host.

#### What does this paper add?

- We demonstrate that phosphoglycerate kinase (PGK) is indeed exposed on pneumococcal surface localising with plasminogen. We provide kinetic evaluation of interaction with human and murine plasminogen, detect two linear plasminogen binding sites and localise the crucial amino acids of PGK contributing to plasminogen binding.
- We present the high-resolution crystal structure of pneumococcal PGK revealing localisation of the plasminogen-binding sites in one surface exposed region of the N-terminal domain and determine the first four kringle domains of plasminogen, named angiostatin, as binding domains required for interaction with PGK.
- Kringle-based interaction between PGK and tPA was predicted by structural modelling and confirmed experimentally by binding studies and plasmin activity assays.

#### Conflicts of interest

None declared.

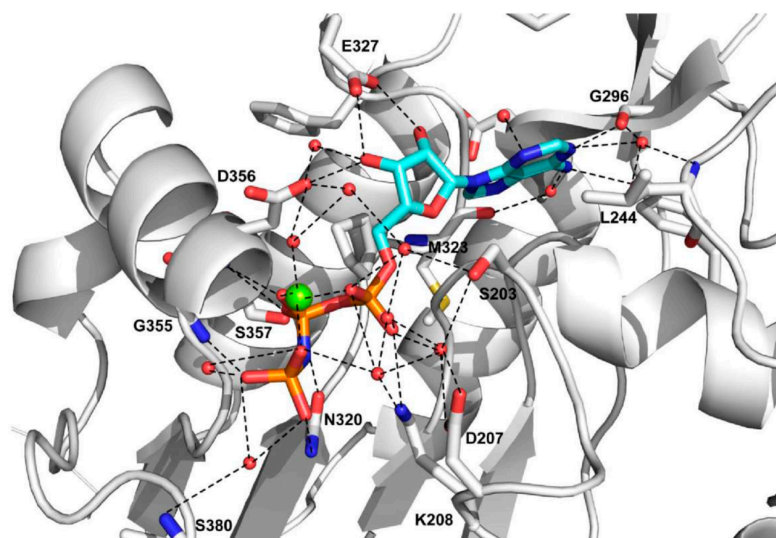
#### References

1. Cartwright K. Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. *Eur J Pediatr* 2002; 161: 188–195.
2. Kadioglu A, Weiser JN, Paton JC, et al. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonisation and disease. *Nat Rev Microbiol* 2008; 6: 288–301.
3. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. *Microbiology* 2006; 152: 295–303.
4. Pérez-Dorado I, Galán-Bartual S, Hermoso JA. Pneumococcal surface proteins: when the whole is greater than the sum of its parts. *Mol Oral Microbiol* 2012; 4: 221–245.
5. Bergmann S, Rohde M, Preissner KT, et al. The nine residue plasminogen-binding motif of the pneumococcal enolase is the major cofactor of plasmin-mediated degradation of extracellular matrix, dissolution of fibrin and transmigration. *Thromb Haemost* 2005; 94: 304–311.
6. Bergmann S, Hammerschmidt S. Fibrinolysis and host response in bacterial infections. *Thromb Haemost* 2007; 98: 512–520.
7. Miyashita C, Wenzel E, Heiden M. Plasminogen: a brief introduction into its biochemistry and function. *Haemostasis* 1988; 18 (Suppl 1): 7–13.
8. Chavakis T, Athanopoulos A, Rhee JS, et al. Angiostatin is a novel anti-inflammatory factor by inhibiting leukocyte recruitment. *Blood* 2005; 105: 1036–1043.
9. Dubocq C, Genoud V, Parborell ME, et al. Impaired clot lysis by rt-PA catalysed mini-plasminogen activation. *Thromb Res* 1997; 86: 505–513.
10. Castellino FJ, Powell JR. Human plasminogen. *Methods Enzymol* 1981; 80: 365–378.
11. Winkler F, Kastenbauer S, Koedel U, et al. Role of the urokinase plasminogen activator system in patients with bacterial meningitis. *Neurology* 2002; 59: 1350–1355.
12. Donofrio FC, Calil AC, Miranda ET, et al. Enolase from *Paracoccidioides brasiliensis*: isolation and identification as a fibronectin-binding protein. *J Med Microbiol* 2009; 58: 706–713.
13. Carneiro CR, Postol E, Nomizo R, et al. Identification of enolase as a laminin-binding protein on the surface of *Staphylococcus aureus*. *Microbes Infect* 2004; 6: 604–608.
14. Bergmann S, Rohde M, Chhatwal GS, et al. alpha-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol Microbiol* 2001; 40: 1273–1287.
15. Hughes MJ, Moore JC, Lane JD, et al. Identification of major outer surface proteins of *Streptococcus agalactiae*. *Infect Immun* 2002; 70: 1254–1259.
16. Bernstein BE, Michels PA, Hol WG. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* 1997; 385: 275–278.
17. Pecorari F, Guilbert C, Minard P, et al. Folding and functional complementation of engineered fragments from yeast phosphoglycerate kinase. *Biochemistry* 1996; 35: 3465–3476.
18. Bergmann S, Lang A, Rohde M, et al. Integrin-linked kinase is required for vitronectin-mediated internalisation of *Streptococcus pneumoniae* by host cells. *J Cell Sci* 2009; 122: 256–267.
19. Frank R, Overwin H. SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. *Methods Mol Biol* 1996; 66: 149–169.
20. Bergmann S, Wild D, Diekmann O, et al. Identification of a novel plasmin(ogen)-binding motif in surface displayed alpha-enolase of *Streptococcus pneumoniae*. *Mol Microbiol* 2003; 49: 411–423.
21. Bernardo-García N, Bartual SG, Fulde M, et al. Crystallisation and preliminary X-ray diffraction analysis of phosphoglycerate kinase from *Streptococcus pneumoniae*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2011; 67: 1285–1289.
22. Leslie AG. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 2006; 62: 48–57.

23. Auerbach G, Huber R, Grättinger M, et al. Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Structure* 1997; 5: 1475-1483.
24. Adams PD, Afonine PV, Bunkoczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; 66: 213-221.
25. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004; 60: 2126-2132.
26. Lyskov S, Gray JJ. The Rosetta Dock server for local protein-protein docking. *Nucleic Acids Res* 2008; 36: W233-238.
27. Duan Y, Wu C, Chowdhury S, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 2003; 24: 1999-2012.
28. Lee MC, Duan Y. Distinguish protein decoys by using a scoring function based on a new AMBER force field. Short molecular dynamics simulations, and the generalized born solvent model. *Proteins* 2004; 55: 620-634.
29. Case DA, Cheatham TE 3rd, Darden T, et al. The Amber biomolecular simulation programs. *J Comput Chem* 2005; 26: 1668-1688.
30. Morreale A, Gil-Redondo R, Ortiz AR. A new implicit solvent model for protein-ligand docking. *Proteins* 2007; 67: 606-616.
31. van der Linden M, Al-Lahham A, Nicklas W, et al. Molecular characterisation of pneumococcal isolates from pets and laboratory animals. *PLoS One* 2009; 4: e8286.
32. Ehinger S, Schubert WD, Bergmann S, et al. Plasmin(ogen)-binding alpha-enolase from *Streptococcus pneumoniae*: crystal structure and evaluation of plasmin(ogen)-binding sites. *J Mol Biol* 2004; 343: 997-1005.
33. Law RH, Caradoc-Davies T, Cowieson N, et al. The X-ray crystal structure of full-length human plasminogen. *Cell Rep* 2012; 1: 185-190.
34. de Vos AM, Ulsch MH, Kelley RE, et al. Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.4-Å resolution. *Biochemistry* 1992; 31: 270-279.
35. Huai Q, Mazar AP, Kuo A, et al. Structure of human urokinase plasminogen activator in complex with its receptor. *Science* 2006; 311: 656-659.
36. Ullberg M, Wiman B, Kronvall G. Binding of tissue-type plasminogen activator (t-PA) to *Neisseria meningitidis* and *Haemophilus influenzae*. *FEMS Immunol Med Microbiol* 1994; 9: 171-177.
37. Abad MC, Arni RK, Grella DK, et al. The X-ray crystallographic structure of the angiogenesis inhibitor angiostatin. *J Mol Biol* 2002; 318: 1009-1017.
38. Kinnby B, Booth NA, Svensater G. Plasminogen binding by oral streptococci from dental plaque and inflammatory lesions. *Microbiology* 2008; 154: 924-931.
39. Boone TJ, Burnham CA, Tyrrell GJ. Binding of group B streptococcal phosphoglycerate kinase to plasminogen and actin. *Microb Pathog* 2011; 51: 255-261.
40. Boone TJ, Tyrrell GJ. Identification of the actin and plasminogen binding regions of group B streptococcal phosphoglycerate kinase. *J Biol Chem* 2012; 287: 29035-29044.
41. Pancholi V, Fischetti VA. alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci. *J Biol Chem* 1998; 273: 14503-14515.
42. Pancholi V. Multifunctional alpha-enolase: its role in diseases. *Cell Mol Life Sci* 2001; 58: 902-920.
43. Fu Q, Figueroa-Losada M, Ploplis VA, et al. The lack of binding of VEK-30, an internal peptide from the group A streptococcal M-like protein, PAM, to murine plasminogen is due to two amino acid replacements in the plasminogen kringle-2 domain. *J Biol Chem* 2008; 283: 1580-1587.
44. Wu TB, Padmanabhan KP, Tulinsky A. The structure of recombinant plasminogen kringle 1 and the fibrin binding site. *Blood Coagul Fibrinolysis* 1994; 5: 157-166.
45. Marti DN, Schaller J, Llinas M. Solution structure and dynamics of the plasminogen kringle 2-AMCHA complex: 3(1)-helix in homologous domains. *Biochemistry* 1999; 38: 15741-15755.
46. Christen MT, Frank P, Schaller J, et al. Human plasminogen kringle 3: solution structure, functional insights, phylogenetic landscape. *Biochemistry* 2010; 49: 7131-7150.
47. Stec B, Yamano A, Whitlow M, et al. Structure of human plasminogen kringle 4 at 1.68 Å and 277 K. A possible structural role of disordered residues. *Acta Crystallogr D Biol Crystallogr* 1997; 53: 169-178.
48. Chang Y, Mochalkin I, McCance SG, et al. Structure and ligand binding determinants of the recombinant kringle 5 domain of human plasminogen. *Biochemistry* 1998; 37: 3258-3271.
49. Trexler M, Vali Z, Patthy L. Structure of the omega-aminocarboxylic acid-binding sites of human plasminogen. Arginine 70 and aspartic acid 56 are essential for binding of ligand by kringle 4. *J Biol Chem* 1982; 257: 7401-7406.
50. Vali Z, Patthy L. The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential for fibrin affinity of the kringle 1 domain. *J Biol Chem* 1984; 259: 13690-13694.
51. Fulde M, Rohde M, Hitzmann A, et al. SCM, a novel M-like protein from *Streptococcus canis*, binds (mini)-plasminogen with high affinity and facilitates bacterial transmigration. *Biochem J* 2011; 434: 523-535.
52. Fulde M, Rohde M, Polok A, et al. Cooperative plasminogen recruitment to the surface of *Streptococcus canis* via M protein and enolase enhances bacterial survival. *mBio* 2013; 12: e00629-12.
53. Sanderson-Smith M, Walker MJ, Ranson M. The Maintenance of High Affinity Plasminogen Binding by Group A Streptococcal Plasminogen-binding M-like Protein Is Mediated by Arginine and Histidine Residues within the a1 and a2 Repeat Domains. *J Biol Chem* 2006; 281: 25965-25971.
54. Parkkinen J, Hacker J, Korhonen TK. Enhancement of tissue plasminogen activator-catalysed plasminogen activation by *Escherichia coli* S fimbriae associated with neonatal septicaemia and meningitis. *Thromb Haemost* 1991; 65: 483-486.
55. DesJardin LE, Boyle MD, Lottenberg R. Group A streptococci bind human plasmin but not other structurally related proteins. *Thromb Res* 1989; 55: 187-193.

## Supplemental Figures to Fulde et al. “Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator” (Thromb Haemost 2014; 111.3)

Figure S1



**Figure S1.** AMP-PNP binding site. PGK residues involved in ligand stabilization are labeled and depicted as sticks. Polar interactions are depicted as dotted lines. Mg cation is represented as green sphere and water molecules as red spheres. The adenine ring is bound in a hydrophobic slot on the surface of the C-domain of the enzyme. The ribose is located above the pyrrolidine ring of Pro322 with the two hydroxyl oxygens of the ribose ring interacting with the carboxylate of Glu327. As in other PGK structures, a  $Mg^{2+}$  ion is coordinated by the three phosphate groups of the nucleotide via three non-bridging oxygen atoms. The remaining three octahedral coordination sites of the divalent metal ion are occupied by two water molecules, one from the symmetry mate, and by Glu108 from the symmetry mate.



## Figure S2

	10	15	55	95	100	105
<i>S. pneumoniae</i>	.... VKDV <b>D</b> L <b>K</b> <b>G</b> <b>K</b> <b>K</b> VL... G <b>R</b> - A I L ... G V T R G A E L <b>E</b> A A I N A L <b>E</b> D ...					
<i>S. aureus</i>	.... V S D L <b>D</b> L <b>K</b> <b>G</b> <b>K</b> T V L... G <b>K</b> - I V L ... P E T R G E K L <b>E</b> A A I K D L K <b>E</b> ...					
<i>Homo sapiens</i>	.... L D K L <b>D</b> V <b>K</b> <b>G</b> <b>K</b> <b>R</b> V V... A <b>K</b> S V V L ... K D C V G P E V <b>E</b> K A C A N P A A ...					
Pig	.... L D K L N V <b>K</b> <b>G</b> <b>K</b> <b>R</b> V V... A <b>K</b> S V V L ... K D C V G P A V <b>E</b> K A C A D P A A ...					
<i>Helicobacter pylori</i>	.... I Q E V <b>E</b> V G H <b>K</b> <b>R</b> V L... A <b>K</b> D I I L ... Q N - - I A Q L K Q A L N E N A P ...					
<i>E. coli</i>	.... M T D L <b>D</b> L A G <b>K</b> <b>R</b> V F... A <b>K</b> - V M V ... P V R L V K D Y L D G - V D V A <b>E</b> ...					
<i>Salmonella</i>	.... M T D L <b>D</b> L A G <b>K</b> <b>R</b> V F... A <b>K</b> - V M V ... P V R L V K D Y L D G - V D V A <b>E</b> ...					
<i>Acinetobacter baumannii</i>	.... M T D L N L A G <b>K</b> <b>R</b> V L... A A - V M V ... E V K L F T D Y L D G - V E V <b>E</b> A ...					
<i>Pseudomonas</i>	.... M T D L <b>D</b> L <b>K</b> <b>G</b> <b>K</b> <b>R</b> V L... A A - V M V ... E V P L L A D Y L D G - V E V K A ...					
<i>Stenotrophomonas</i>	.... M T D L <b>D</b> L S G <b>K</b> <b>R</b> V L... A A - V M V ... E V P L V R D W V D G - V D V Q P ...					
<i>Neisseria</i>	.... L T E Q N V Q G <b>K</b> T V L... A S - V I V ... D V K V L N D W R E N K P A L N A ...					
<i>Desulfurococcus</i>	.... V R D V N I S G <b>K</b> <b>K</b> V F... Q P A L V L ... P D I I G P R A L E E I N N L K P ...					

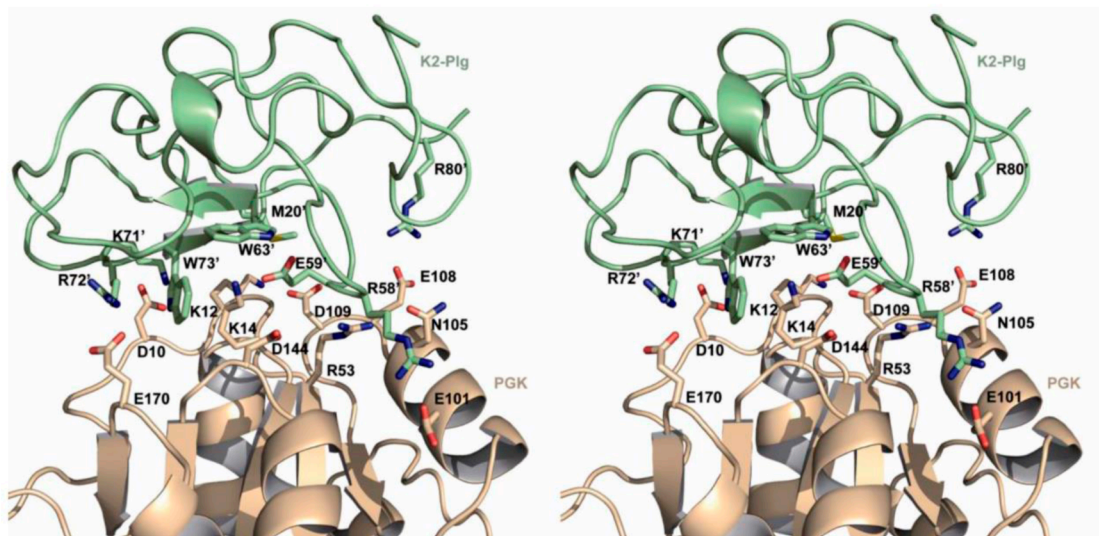
  

	145	170
<i>S. pneumoniae</i>	... S L G <b>D</b> G ... V <b>E</b> <b>K</b> A V ...	
<i>S. aureus</i>	... S L G <b>D</b> - ... L <b>E</b> T A A ...	
<i>Homo sapiens</i>	... S L S K L ... L P Q K A ...	
Pig	... S L S A L ... L P <b>K</b> K A ...	
<i>Helicobacter pylori</i>	... S L C <b>D</b> - ... A P I K V ...	
<i>E. coli</i>	... A L C <b>D</b> - ... A <b>D</b> V A C ...	
<i>Salmonella</i>	... A L C <b>D</b> - ... A <b>D</b> V A C ...	
<i>Acinetobacter baumannii</i>	... A L C <b>D</b> - ... A P V A A ...	
<i>Pseudomonas</i>	... A L C <b>D</b> - ... A K V A A ...	
<i>Stenotrophomonas</i>	... A L C <b>D</b> - ... A P V A A ...	
<i>Neisseria</i>	... S L C <b>D</b> - ... A P V A C ...	
<i>Desulfurococcus</i>	... A V F N - ... M P S C M ...	

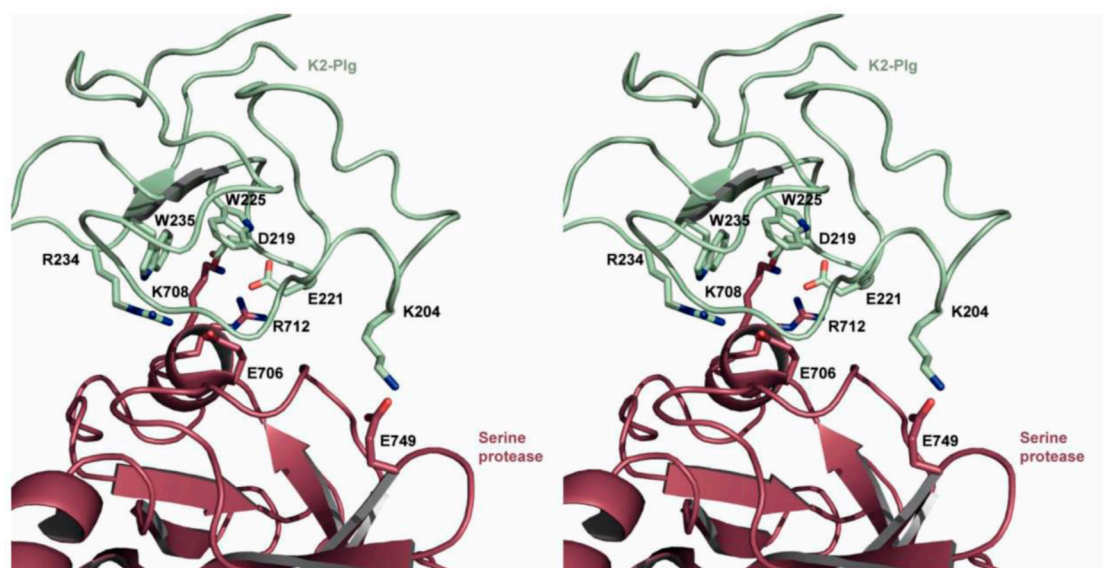
**Figure S2.** Sequence alignment among PGK from different species. Amino acids from the plasminogen binding site are in bold, acidic residues are depicted in red and basic residues in blue, conservative mutations are depicted in light colors.

## Figure S3

A



B



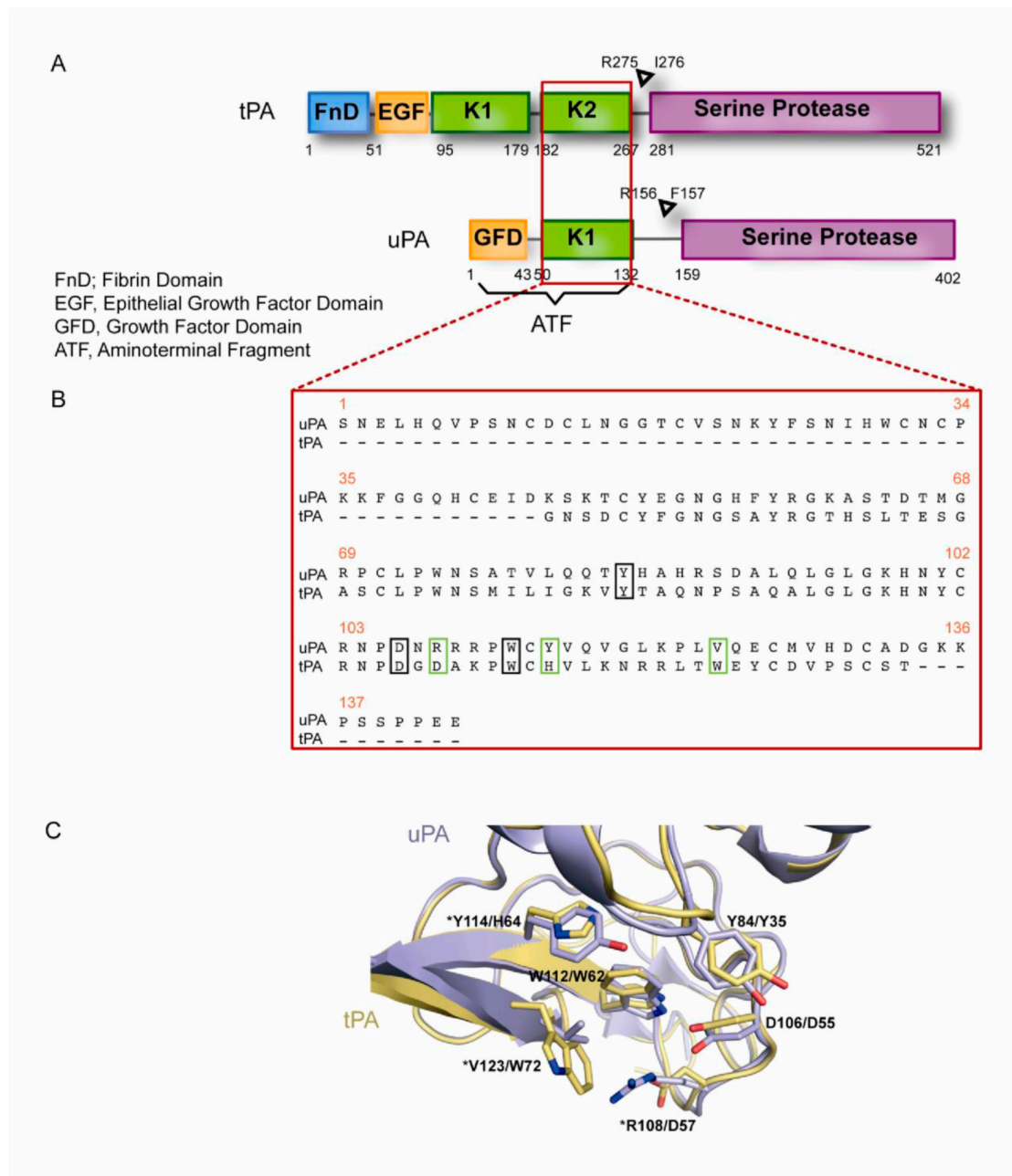
**Figure S3.** Comparison of the predicted interactions between pneumococcal PGK and human K2 and observed interactions between K2 and Serine protease (SP) domains in crystal structure of full-length plasminogen. **A.** Stereoview representation of K2 (green) interacting with PGK (brown). **B.** Stereoview representation of K2 domain (green) and SP domain (pink) in human plasminogen. In both cases, K2 lysine-binding sites present a lysine residue inserted in their Trp hydrophobic establishing a salt bridge interaction with an acidic residue from K2. Other salt bridge and polar interactions are produced around this central interaction.

Figure S4

Position	10	20	30	40	50	60	70	80
<i>S. pneumoniae</i>	MAKLTVKDV	DLKGKKVLVRVDFNVP	LKDGVITNDNRI	TAALPTIKYI	IEQGGRILF	SHLGRVKEE	SDK	AGKSLAPVAAD
<i>S. agalactiae</i>	MAKLTVKDV	DLKGKKVLVRVDFNVP	LKDGVITNDNRI	TAALPTIKYI	IEQGGRILF	SHLGRVKEE	ADKEGKSLAPVAAD	
<i>S. anginosus</i>	MAKLTVKDV	DLKGKKVLVRVDFNVP	LKDGVITNDNRI	SAALPTIKYI	IEHGGRILF	SHLGRVKEE	ADKEGKSLAPVAAD	
<i>S. oralis</i>	MAKLTVKDV	DLKGKKVLVRVDFNVP	VKDGVITNDNRI	TAALPTIKYI	IEQGGRILF	SHLGRVKEE	ADKEGKSLAPVAAD	
Position	90	100	110	120	130	140	150	160
<i>S. pneumoniae</i>	LA	AKLGQDVVF	-PGVTRGAE	LEAA	INALE	DGQVLLVENTR	YEDVDG	KKESKND
<i>S. agalactiae</i>	LA	AKLGQDVVF	-PGVTRGAK	LEEA	INALE	DGQVLLVENTR	FEDVDG	KKESKND
<i>S. anginosus</i>	LA	AKLGQDVVF	-PGVTRGAE	LEAA	INALE	DGQVLLVENTR	FEDVDG	KKESKND
<i>S. oralis</i>	LA	AKLGQDVVF	-PGVTRGAE	LEAA	INALE	DGQVLLVENTR	FEDVDG	KKESKND
Position	170	180	190	200	210	220	230	240
<i>S. pneumoniae</i>	ASNVGISANV	EKAVAGFLLENE	IAYIQEAVE	TPERPFVAILGGS	KVSDKIGV	ENLLEKADK	VLIGGGMTY	TFYKAQGGIE
<i>S. agalactiae</i>	ASNVGISANV	EKAVAGFLLENE	IAYIQEAVE	TPERPFVAILGGS	KVSDKIGV	ENLLEKADK	VLIGGGMTY	TFYKAQGGIE
<i>S. anginosus</i>	ASNVGISANV	EKAVAGFLLENE	IAYIQEAVE	TPERPFVAILGGS	KVSDKIGV	ENLLEKADK	VLIGGGMTY	TFYKAQGGIE
<i>S. oralis</i>	ASNVGISANV	EKAVAGFLLENE	IAYIQEAVE	TPERPFVAILGGS	KVSDKIGV	ENLLEKADK	VLIGGGMTY	TFYKAQGGIE
Position	250	260	270	280	290	300	310	320
<i>S. pneumoniae</i>	IGNSLVEEDKL	DVAKALLEK	ANGKLI	LPVDSKEANAF	AGYTEVR	RDTEGEAV	SEGFLGLD	IGPKSI
<i>S. agalactiae</i>	IGNSLVEEDKL	DVAKALLEK	ANGKLI	LPVDSKEANAF	AGYTEVR	RDTEGEAV	SEGFLGLD	IGPKSI
<i>S. anginosus</i>	IGNSLVEEDKL	DVAKALLEK	ANGKLI	LPVDSKEANAF	ADYTEVR	KDTEGEAV	DPGFLGLD	IGPKSI
<i>S. oralis</i>	IGNSLVEEDKL	DVAKALLEK	ANGKLI	LPVDSKEANAF	ADYTEVR	KDTEGEAV	DPGFLGLD	IGPKSI
Position	330	340	350	360	370	380	390	
<i>S. pneumoniae</i>	NGPMGVFENPDF	QAGTIGVMDA	IVKQPGVKS	IIGGGDSAAAA	INLGRADK	FSWISTGGG	ASMELEGG	VLPLGLAALTEK
<i>S. agalactiae</i>	NGPMGVFENPDF	QAGTIGVMDA	IVKQPGVKS	IIGGGDSAAAA	INLGRADK	FSWISTGGG	ASMELEGG	VLPLGLAALTEK
<i>S. anginosus</i>	NGPMGVFENPDF	QAGTIGVMDA	IVKQPGVKS	IIGGGDSAAAA	INLGRADK	FSWISTGGG	ASMELEGG	VLPLGLAALTEK
<i>S. oralis</i>	NGPMGVFENPDF	QAGTIGVMDA	IVKQPGVKS	IIGGGDSAAAA	INLGRADK	FSWISTGGG	ASMELEGG	VLPLGLAALTEK

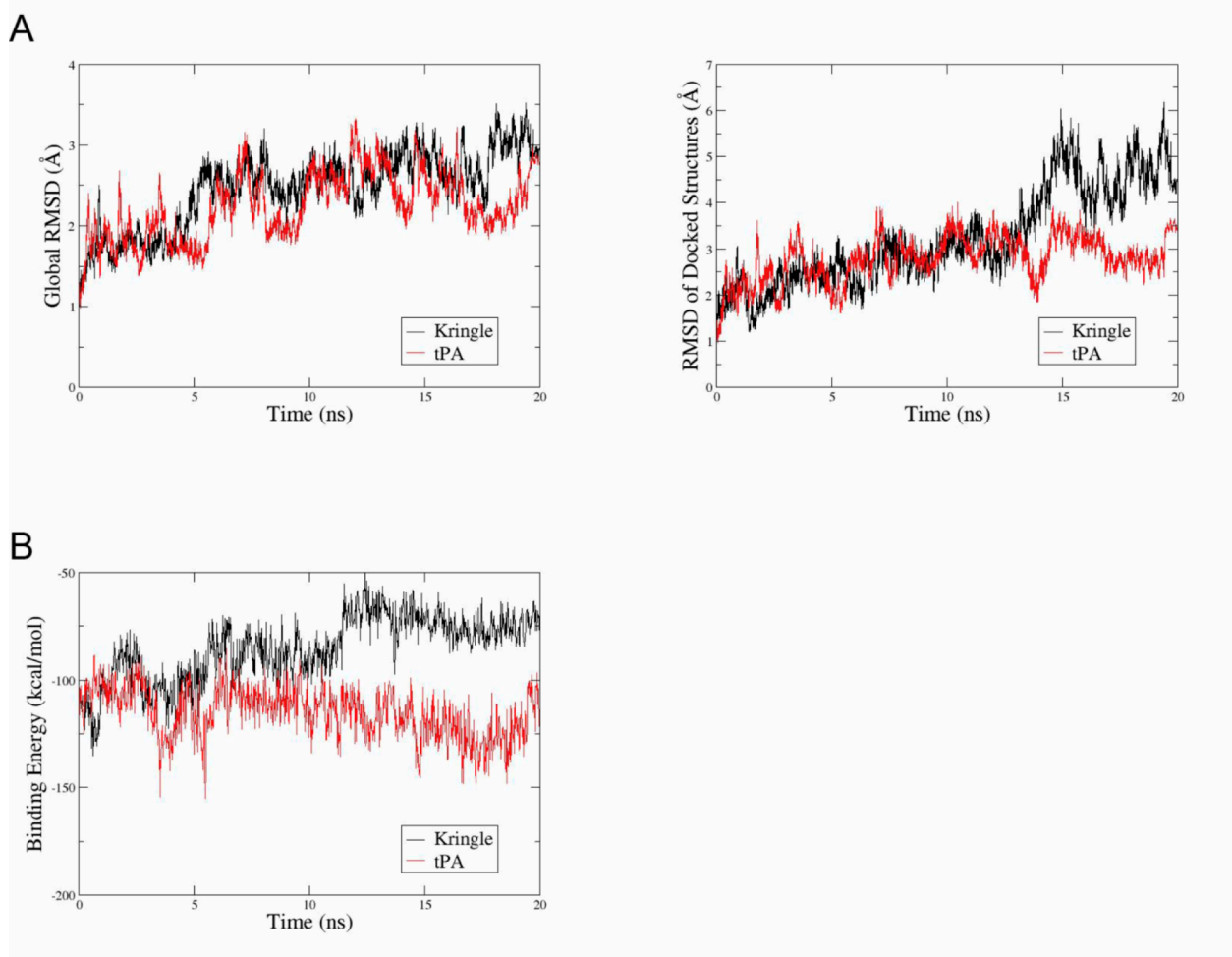
**Figure S4.** Amino acid-sequence alignment of PGK proteins expressed by *Streptococcus pneumoniae* serotype 2 (D39; AccNo.: YP\_815953), *Streptococcus agalactiae* (ATCC13813; Acc.No.: ZP\_08649963.1), *Streptococcus anginosus* (ZP\_07864656.1) and *Streptococcus oralis* (ATCC35037; ZP\_06610927.1). Differences in amino acid sequence are marked in red, plasminogen binding sequence of *S. agalactiae* is highlighted in green, presence of proline at amino acid position 133 is marked in blue.



**Figure S5**

**Figure S5.** Structural and sequence analysis comparison between tPA and uPA. **A.** Schematic model of tPA and uPA domain structure. Triangles indicate the cleavage site. Red box shows the compared kringle domains from both activators. **B.** Sequence alignment between uPA and tPA. Boxes indicate residues involved in PGK binding, green boxes show residues not conserved in uPA. **C.** tPA and uPA superimposition. Residues from uPA (blue) and tPA (yellow) involved in PGK binding are in sticks, asterisk marked uPA residues implied in the lacking of binding.

## Figure S6



**Figure S6.** Root-mean squared deviation (RMSD) plots of kringle domain and binding energies of interaction between PGK and kringle domain 2. **A.** Plot showing the RMSD fluctuations for the final kringle domain (black line) and tPA (red line) attached to the PGK for the 10ns MD simulations. **B.** Plot showing binding energies between PGK and K2 kringle domain (black line) and between PGK and tPA (red line). In both cases interaction between K2 and tPA with PGK are similar.

#### 4.2.1. DISCUSIÓN

En este trabajo se ha determinado que la proteína PGK forma parte del grupo de proteínas que llevan a cabo funciones diferentes a las del metabolismo celular y que se localiza también en la superficie celular interaccionando con las proteínas de la fibrinólisis. Se ha demostrado la co-localización de la PGK y el PLG, confirmando así que la PGK es una proteína que se une a PLG.

El análisis estructural de la proteína PGK indicó que el sitio de unión a PLG se encuentra localizado en el extremo N-terminal en la cara opuesta al sitio catalítico, por lo que no afectaría a la actividad enzimática. Además, se ha determinado que la unión de la PGK al PLG no genera cambios conformacionales.

El PLG interacciona con otras proteínas a través de los sitios de unión a lisina de sus dominios *kringle*. Estos sitios de unión están constituidos por dos triptófanos y una serie de residuos cargados que los rodean. La estructura de la PGK demostró que existe una zona rica en residuos de lisina expuestos que median en la interacción con el dominio *kringle* del PLG.

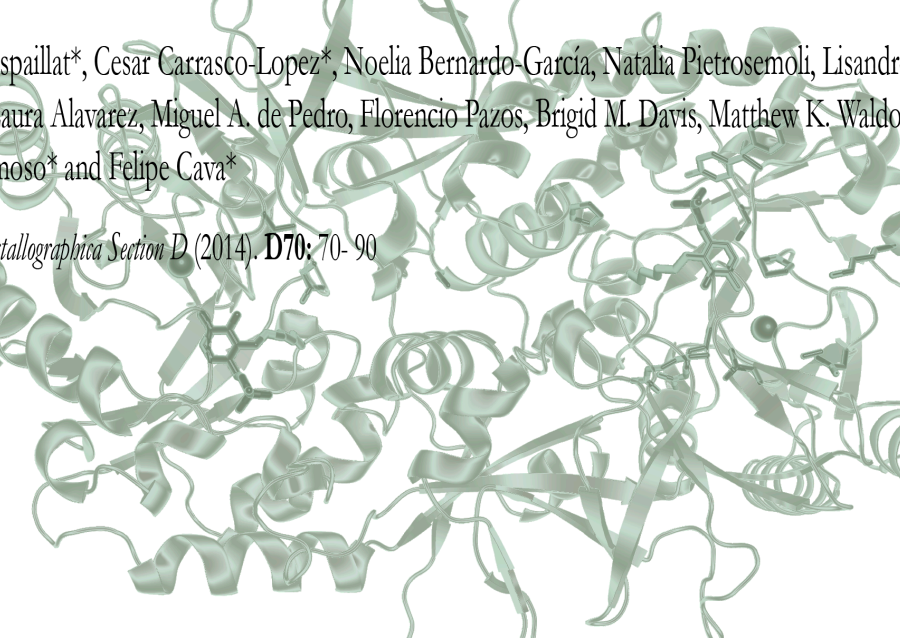
Se ha realizado un modelo computacional de la interacción entre la PGK y el *kringle* 2 (K2) de la angiostatina. Apoyados en este modelo hemos propuesto que otras proteínas que contienen dominios *kringle*, como son los activadores de PLG (tPA y uPA), podrían interaccionar también con la PGK. A través de diferentes ensayos bioquímicos se ha confirmado que PGK es capaz de unirse a tPA pero no a uPA. En este trabajo se demuestra por primera vez la unión de *S. pneumoniae* al activador de PLG tPA. Además, hemos explicado por qué no interacciona con el uPA.

La unión de la PGK a los dominios *kringle* del PLG permite que éste se encuentre en una conformación abierta promoviendo así su activación. Se ha demostrado que el reclutamiento de la plasmina en la superficie del neumococo facilita la degradación de fibrina, trombina y la matriz extracelular, favoreciendo la diseminación a través de los tejidos y la entrada en el sistema vascular. La unión de la PGK al PLG promueve la conversión del PLG en plasmina a través de la tPA, una propiedad única de la PGK es la capacidad de unir tPA, lo cual también fomenta la activación del PLG y en consecuencia la degradación de la fibrina. La inmovilización tanto del PLG como del activador tPA en la superficie bacteriana acelera la conversión enzimática del PLG a través del tPA en el lugar donde se está produciendo la infección por neumococo, facilitando así la dispersión del patógeno en el hospedador.

Considerando que la fibrinólisis participa en los procesos inflamatorios complejos, nuestros resultados describen un nuevo escenario en el cual el neumococo podría revertir la maquinaria de la fibrinólisis durante la fase inicial de colonización aumentando su capacidad para diseminarse a través de los tejidos y entrando en el sistema vascular.

Las conclusiones de este trabajo se resumen a continuación:

- ✓ Se ha resuelto la estructura cristalográfica de la proteína PGK, que presenta un sitio de unión a plasminógeno en la cara opuesta a donde se encuentra el sitio activo de la enzima.
- ✓ Se ha demostrado que la PGK es capaz de unir plasminógeno y su activador tPA.
- ✓ Se ha propuesto un mecanismo de acción del neumococo a través del cual es capaz de reclutar PLG y tPA en la superficie aumentando así su diseminación a través de los tejidos y entrada al sistema circulatorio.

[illegible]

# amino-acid racemases

Akbar Espaillat\*, Cesar Carrasco-Lopez\*, Noelia Bernardo-García, Natalia Pietrosemoli, Lisandro H. Otero, Laura Alavarez, Miguel A. de Pedro, Florencio Pazos, Brigid M. Davis, Matthew K. Waldor, Juan A., Hermoso\* and Felipe Cava\*

*Acta Crystallographica Section D* (2014). **D70**: 70-90

Acta Crystallographica Section D (2014). D70: 70-90



### 4.3. INTRODUCCIÓN A LA RACEMASA DE AMPLIO ESPECTRO BsrV Y APORTACIÓN DEL AUTOR

En la naturaleza la mayoría de los aminoácidos D (DAA) se encuentran en las paredes bacterianas, D-Ala y D-Glu son los constituyentes de los péptidos que forman las cadenas peptídicas del peptidoglicano (PG). Estos DAA se producen a partir de las formas L a través de las enzimas alanina y glutamato racemasas (Alr1 y Mur1). La enzima Alr1 ha sido muy estudiada y está muy bien caracterizada, es una proteína citoplasmática que necesita del coenzima piridoxal fosfato (PLP) como cofactor. Mediante estudios cristalográficos de la AlaR de diversos organismos se han establecido las características estructurales que determinan su mono-especificidad.

Recientemente, se ha demostrado que diversas bacterias producen y liberan al medio grandes cantidades de DAA distintos de los componentes canónicos de la pared celular (D-Ala y D-Glu). En *Vibrio cholerae*, la producción de estos DAA no canónicos (NCDAA) es dependiente de la racemasa BsrV, una enzima periplásmica de amplio espectro con una homología de secuencia muy alta con la alanina racemasa de *V. cholerae* (AlrV). En *V. cholerae*, la presencia de NCDAA regula negativamente el metabolismo celular, a través de su incorporación en el PG. Los NCDAA también afectan a la estabilidad del biofilm, a la virulencia y a la esporulación. Se ha relacionado la producción de NCDAA con la competición de diferentes bacterias por establecerse en un hábitat. Dada la importancia de los NCDAA como agentes proveedores de propiedades adaptativas, el estudio de las racemasas multiespecíficas es crítico para entender los procesos que controlan, así como para desarrollar nuevos fármacos para combatir las bacterias patógenas.

En este trabajo, se ha investigado la capacidad de la BsrV para racemizar una amplia variedad de sustratos usando una combinación de técnicas estructurales, bioquímicas, de ingeniería de proteínas y bioinformáticas. Además, se ha determinado la huella molecular de los residuos que determina la multiespecificidad de la BsrV a través de la cual se ha podido identificar una familia de racemasas multiespecíficas de las que BsrV sería su representante.

A través de este trabajo nos planteamos los siguientes objetivos:

- ✓ Se ha resuelto la estructura cristalográfica de la proteína BsrV a muy buena resolución atómica. Así como la de la racemasa de amplio espectro de *Aeromonas hydrophila*, Bsr<sub>Ah</sub>.
- ✓ Se ha demostrado que la BsrV presenta un mecanismo de reacción diferente al de las racemasas mono-específicas, con unos determinantes estructurales únicos.
- ✓ Se ha identificado la huella molecular que diferencia las racemasas de amplio espectro de aquellas que son mono-específicas.

En esta publicación, la autora de esta Tesis ha realizado parte de los ensayos de cristalización, resolución estructural, el análisis estructural y el desarrollo de los modelos de *docking* que se proponen en el trabajo de investigación.



**Structural basis for the broad specificity of a new family of amino-acid racemases.**

Akbar Espaillat\*, Cesar Carrasco-Lopez\*, Noelia Bernardo-García, Natalia Pietrosevoli, Lisandro H. Otero, Laura Alvarez, Miguel A. de Pedro, Florencio Pazos, Brigid M. Davis, Matthew K. Waldor, Juan A. Hermoso\* y Felipe Cava\*

Acta Crystallographica Section D (2014) D70: 70- 90



Acta Crystallographica Section D

Biological  
Crystallography

ISSN 1399-0047

## Structural basis for the broad specificity of a new family of amino-acid racemases

Akbar Espailat,<sup>a,‡</sup> Cesar Carrasco-Lopez,<sup>b,‡</sup> Noelia Bernardo-García,<sup>b</sup> Natalia Pietrosevoli,<sup>c</sup> Lisandro H. Otero,<sup>b</sup> Laura Alvarez,<sup>a</sup> Miguel A. de Pedro,<sup>a</sup> Florencio Pazos,<sup>c</sup> Brigid M. Davis,<sup>d</sup> Matthew K. Waldor,<sup>d</sup> Juan A. Hermoso<sup>b,\*</sup> and Felipe Cava<sup>a,§</sup><sup>a</sup>Centro de Biología Molecular 'Severo Ochoa', Universidad Autónoma de Madrid–Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain, <sup>b</sup>Department of Crystallography and Structural Biology, Instituto de Química-Física 'Rocasolano'–CSIC, 28006 Madrid, Spain, <sup>c</sup>Centro Nacional de Biotecnología–CSIC, 28049 Madrid, Spain, and <sup>d</sup>Division of Infectious Diseases, Brigham and Women's Hospital and Department of Microbiology and Immunobiology, Harvard Medical School and HHMI, Boston, MA 02115, USA<sup>‡</sup> These authors contributed equally.<sup>§</sup> Current address: Department of Molecular Biology and Laboratory for Molecular Infection Medicine Sweden, Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden.

Correspondence e-mail: xjuan@iqfr.csic.es, felipe.cava@molbiol.umu.se

Broad-spectrum amino-acid racemases (Bsr) enable bacteria to generate noncanonical D-amino acids, the roles of which in microbial physiology, including the modulation of cell-wall structure and the dissolution of biofilms, are just beginning to be appreciated. Here, extensive crystallographic, mutational, biochemical and bioinformatic studies were used to define the molecular features of the racemase BsrV that enable this enzyme to accommodate more diverse substrates than the related PLP-dependent alanine racemases. Conserved residues were identified that distinguish BsrV and a newly defined family of broad-spectrum racemases from alanine racemases, and these residues were found to be key mediators of the multispecificity of BsrV. Finally, the structural analysis of an additional Bsr that was identified in the bioinformatic analysis confirmed that the distinguishing features of BsrV are conserved among Bsr family members.

## 1. Introduction

In nature, the majority of D-amino acids (DAA) are found in bacterial cell walls, where D-Ala and D-Glu are key constituents of the short peptides that cross-link the glycan chains within the peptidoglycan (PG) polymer (Vollmer *et al.*, 2008). These DAA are produced from their corresponding L-enantiomers by highly specific alanine and glutamate amino-acid racemases (Alr and MurI), which have distinct structural and catalytic properties (Cava, Lam *et al.*, 2011). Alr (EC 5.1.1.1), which is the best characterized of these enzymes, is a cytoplasmic protein that requires the ubiquitous coenzyme pyroxidol-5-phosphate (PLP) as a cofactor (Richard *et al.*, 2009). Crystallographic studies of Alr from diverse organisms have enabled the identification of the structural determinants and spatial constraints that account for its mono-specificity (Couñago *et al.*, 2009; LeMagueres *et al.*, 2003, 2005; Noda *et al.*, 2004; Shaw *et al.*, 1997; Strych & Benedik, 2002; Wu *et al.*, 2008).

Recently, we demonstrated that diverse bacteria produce and release to the extracellular medium high amounts of DAA that are not canonical components of the cell wall (*e.g.* D-Ala and D-Glu; Cava, de Pedro *et al.*, 2011; Lam *et al.*, 2009). In *Vibrio cholerae*, the production of these noncanonical DAA (NCDAA) is dependent upon BsrV (VC1312), a periplasmic broad-spectrum racemase (Bsr) with high sequence homology (*E*-value of  $1 \times 10^{-25}$ , 28% identity) to the primary alanine racemase of *V. cholerae*, AlrV (VC0372; Fig. 1a and Supplementary Fig. S1a<sup>1</sup>; Lam *et al.*, 2009). A subset of Alr-related racemases from other organisms have also been reported to accept multiple amino-acid substrates (Kino *et al.*, 2007; Lam

Received 2 August 2013

Accepted 5 September 2013

**PDB References:** BsrV, 4beu;  
 $\Delta$ Cl-BsrV, 4beq; Bsr<sub>Alr</sub>, 4bf5;  
 Alr<sub>Alr</sub>, 4bhy

## research papers

*et al.*, 2009; Matsui & Oikawa, 2010; Wu *et al.*, 2008). In *V. cholerae*, the presence of NCDAA negatively regulates cell-wall metabolism, at least in part owing to their incorporation into PG (Cava, de Pedro *et al.*, 2011; Cava, Lam *et al.*, 2011; Lam *et al.*, 2009). NCDAA have also been reported to influence biofilm stability (Kolodkin-Gal *et al.*, 2010), virulence (Anfora *et al.*, 2007) and sporulation (O'Connor & Zusman, 1997). Susceptibility to NCDAA seems to be more widespread than production of NCDAA, raising the possibility that some bacteria may produce and release NCDAA in order to influence nonproducing species within the same habitat (Cava, de Pedro *et al.*, 2011; Horcajo *et al.*, 2012). Given the important

adaptative properties provided by NCDAA, a detailed investigation of multi-specific racemases will be critical to understand NCDAA-controlled processes and for the development of new drugs to combat pathogenic bacteria (Conti *et al.*, 2011).

Here, we explore the capacity of BsrV to utilize a wide variety of substrates using a combination of structural, biochemical, mutational and bioinformatic approaches. Finally, we also define and experimentally validate a molecular footprint of conserved residues that mediate the multi-specificity of BsrV and use this signature sequence to identify a large family of BsrV-like multi-specific racemases in bacteria.

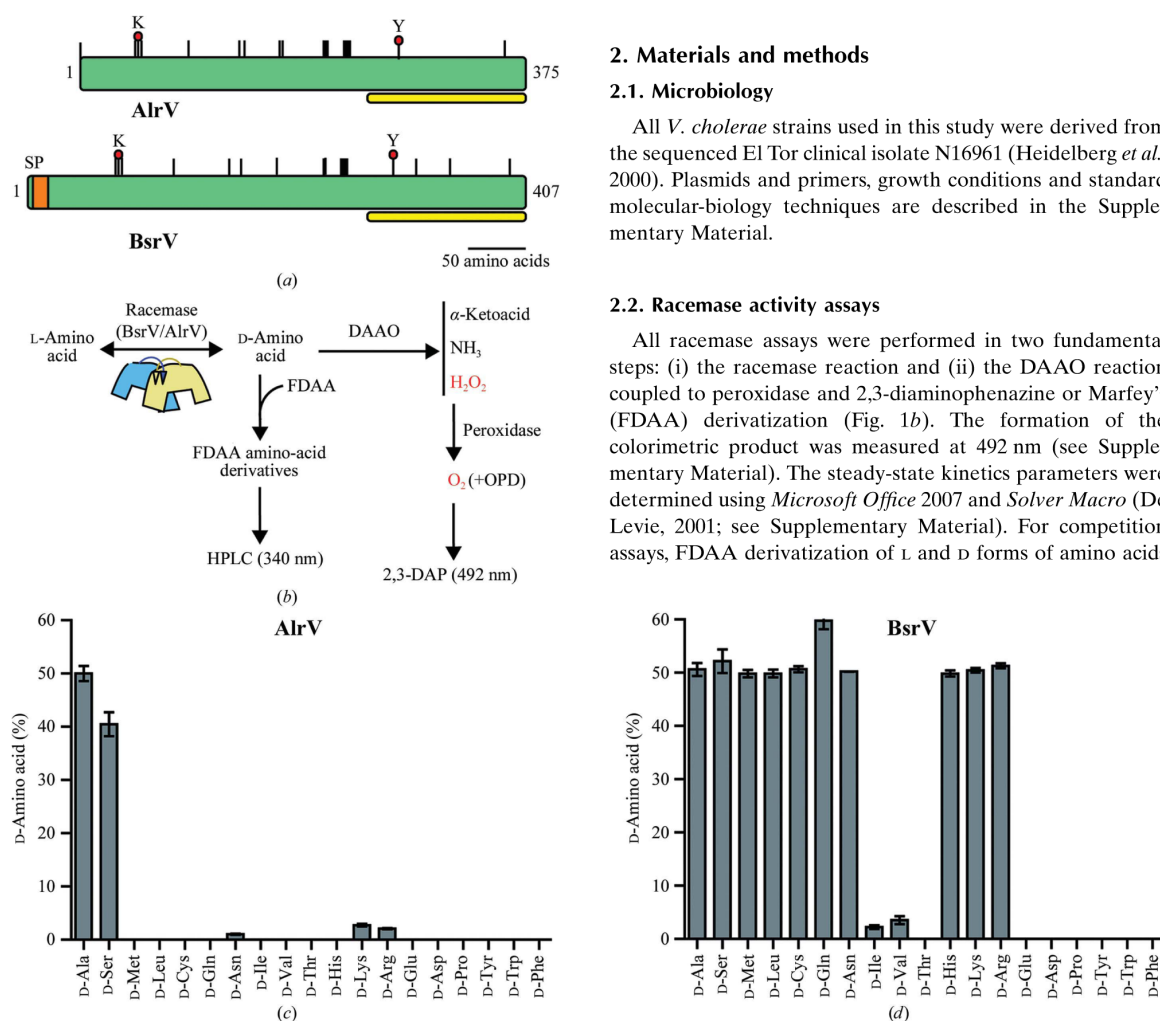
## 2. Materials and methods

## 2.1. Microbiology

All *V. cholerae* strains used in this study were derived from the sequenced El Tor clinical isolate N16961 (Heidelberg *et al.*, 2000). Plasmids and primers, growth conditions and standard molecular-biology techniques are described in the Supplementary Material.

## 2.2. Racemase activity assays

All racemase assays were performed in two fundamental steps: (i) the racemase reaction and (ii) the DAAO reaction coupled to peroxidase and 2,3-diaminophenazine or Marfey's (FDAA) derivatization (Fig. 1b). The formation of the colorimetric product was measured at 492 nm (see Supplementary Material). The steady-state kinetics parameters were determined using *Microsoft Office 2007* and *Solver Macro* (De Levie, 2001; see Supplementary Material). For competition assays, FDAA derivatization of L and D forms of amino acids



**Figure 1** Biochemical comparison of AlrV and BsrV from *V. cholerae*. (a) Predicted protein features of the monospecific alanine racemase (AlrV) and broad-spectrum racemase (BsrV) from *V. cholerae* generated with the NCBI protein-analysis package. Dimerization motifs are in yellow. SP, signal peptide. Red circles pinpoint the catalytic Lys and Tyr residues. Vertical lines correspond to conserved functional residues (PLP binding, substrate binding, catalytic site and dimeric interface). (b) Schematic of the different methods used to assess amino-acid racemase activities (see §2). (c, d) D-Amino-acid accumulation in the presence of AlrV (c) and BsrV (d) after 90 min reactions with various substrates revealed through Marfey's derivatization. The results in (c) and (d) are the means  $\pm$  SD of triplicates from two independent experiments. See also Supplementary Fig. S1.

**Table 1**

Kinetic parameters ( $K_m$ ,  $k_{cat}$  and  $k_{cat}/K_m$ ) of AlrV and BsrV for various amounts (1.25–60 mM) of L-amino-acid substrates were calculated using data obtained with 0.47  $\mu$ M purified racemase (see §2).

All kinetic constants must be considered as apparent values because of the impossibility of calculating initial enzyme velocities. The results are means  $\pm$  SD of triplicates from two independent experiments. BsrV reaction kinetics for His and Cys were not determined (ND).

	$K_m$ (mM)	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ ( $mM^{-1} s^{-1}$ )
<b>BrsV</b>			
Ala	11 $\pm$ 2	1.68 $\pm$ 0.11	15.3 $\times 10^{-2}$
Ser	28 $\pm$ 3	4.68 $\pm$ 0.21	16.7 $\times 10^{-2}$
Met	11 $\pm$ 2	2.94 $\pm$ 0.04	26.7 $\times 10^{-2}$
Leu	30 $\pm$ 7	2.30 $\pm$ 0.28	7.7 $\times 10^{-2}$
Cys	ND	ND	ND
Gln	22 $\pm$ 3	2.59 $\pm$ 0.16	11.8 $\times 10^{-2}$
Asn	15 $\pm$ 4	0.05 $\pm$ 0.004	0.3 $\times 10^{-2}$
His	ND	ND	ND
Lys	9 $\pm$ 1	4.76 $\pm$ 0.11	52.9
Arg	18 $\pm$ 3	5.09 $\pm$ 0.11	28.3
<b>AlrV</b>			
Ala	1.5 $\pm$ 0.6	1.42 $\pm$ 0.003	94.5
Ser	17 $\pm$ 8	1.16 $\pm$ 0.0002	6.8

and HPLC analysis was performed (see Supplementary Material).

### 2.3. Structural determination

Crystallization of BsrV, an R<sub>173</sub>N<sub>174</sub>/AA point mutant ( $\Delta$ Cl-BsrV), *Aeromonas hydrophila* Alr3 (a putative Bsr, renamed Bsr<sub>Ah</sub>) and the putative primary alanine racemase from *A. hydrophila* (Alr<sub>Ah</sub>) was performed using a high-throughput NanoDrop ExtY robot (Innovadyne Technologies Inc.) and the commercial Qiagen screens The JCSG+ Suite and The PACT Suite and the Hampton Research screens Index, Crystal Screen and Crystal Screen 2 following the sitting-drop vapour-diffusion method (see Supplementary Material). X-ray data collection was performed at a synchrotron-radiation facility on beamline ID29 at ESRF, Grenoble, France or on the X06SA beamline at SLS, Villigen, Switzerland (see Supplementary Material). Data sets were collected using a PILATUS 6M detector and were processed using *XDS* (Kabsch, 2010) and scaled using *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011). The structure was solved by the molecular-replacement method with *MOLREP* (Vagin & Teplyakov, 2010) from the *CCP4* suite and was refined with *PHENIX* (Adams *et al.*, 2010) and manually improved using *Coot* (Emsley & Cowtan, 2004; see Supplementary Material). The stereochemistry of the models was verified using *MolProbity* (Chen *et al.*, 2010).

### 2.4. Bioinformatic analyses

The identification of new putative broad-spectrum racemases and the residues responsible for substrate specificity was performed by searching with *BLAST* (Altschul *et al.*, 1997) against a nonredundant protein database and filtering, selecting only those mapped onto the UniProt database. The protein data set was aligned using *MUSCLE* (Edgar, 2004; v.3.8.31). The resulting multiple sequence alignment was

filtered using the *Jalview* (Waterhouse *et al.*, 2009) tool (v.2.7). The identification of specificity-determining positions (SDPs) was carried out using the *JDet* package (see Supplementary Material).

### 2.5. Accession numbers

The atomic coordinates and structure factors for BsrV,  $\Delta$ Cl-BsrV mutant, Bsr<sub>Ah</sub> and Alr<sub>Ah</sub> (PDB entries 4beu, 4beq, 4bf5 and 4bhy, respectively) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, New Jersey, USA (<http://www.rcsb.org/>).

## 3. Results

### 3.1. BsrV is a promiscuous paralogue of the *V. cholerae* alanine racemase AlrV

Comparisons of the substrate specificities of BsrV and AlrV, the presumed alanine-specific racemase from *V. cholerae*, revealed that BsrV is able to utilize a much broader range of substrates (Figs. 1*b*, 1*c* and 1*d*). Unlike AlrV, which only catalyzes the interconversion of L-Ala and L-Ser and their corresponding D forms (Fig. 1*c* and Supplementary Fig. S1*b*), BsrV reversibly racemizes ten of the 19 natural chiral amino acids known, including both non- $\beta$ -branching aliphatic amino acids (Ala, Leu, Met, Ser, Cys, Gln and Asn) and positively charged amino acids (His, Lys and Arg) (Fig. 1*d* and Supplementary Figs. S1*c* and S1*d*). Additionally, it catalyzes the racemization of several amino acids that are not typically incorporated into proteins (Supplementary Fig. S1*e*). However, BsrV did not alter the chirality of negatively charged (*i.e.* Glu and Asp) or aromatic (*i.e.* Tyr, Trp and Phe) amino acids and displayed minimal activity towards  $\beta$ -branched aliphatic (*i.e.* Ile, Val and Thr) potential substrates (Fig. 1*d*).

### 3.2. Reaction kinetics of BsrV and AlrV

Kinetic analyses, using individual and pooled substrates, were also performed in order to better characterize the activities of BsrV and AlrV. Most analyses were performed using the His-tagged mature proteins used in structural studies because control assays performed with untagged mature BsrV showed that the affinity tag did not alter the activity of the enzyme (Supplementary Fig. S1*f*). The kinetic analyses revealed that the catalytic efficiency of BsrV (as reflected by  $k_{cat}/K_m$ ) varies depending upon the amino-acid substrate and suggest that it acts more readily upon lysine, arginine and methionine than upon other substrates (Table 1). Assays using pooled substrates largely confirmed this idea: BsrV incubated with its ten potential L-amino-acid substrates efficiently formed D-Arg and D-Lys, but produced lower amounts of other DAA than when they were assayed singly (Supplementary Figs. S1*g* and S1*h*). Consequently, the high production of D-Met and D-Leu by *V. cholerae* *in vivo* (the principal products of BsrV; Lam *et al.*, 2009) is likely to reflect the relatively high availability of the corresponding L-amino acids.

## research papers

Comparative analyses of BsrV and AlrV reaction kinetics revealed that the average  $k_{\text{cat}}$  of BsrV is higher than that of AlrV. However, the  $K_m$  values for BsrV were also around tenfold higher than the  $K_m$  of AlrV for alanine, its preferred substrate, and consequently  $k_{\text{cat}}/K_m$  for alanine racemization by AlrV is higher than that of BsrV for its substrate amino acids (Table 1). These facts fit well with the proposal that the development of catalytic multi-specificity is often accompanied by a reduction in kinetic performance (Babtie *et al.*, 2010; Khersonsky *et al.*, 2006).

### 3.3. The crystal structure reveals unique structural features in BsrV

Purified BsrV was crystallized and its structure was solved at atomic resolution (1.15 Å; Table 2). As observed for monospecific Alrs, BsrV folds as a globular homodimeric enzyme in which both monomers participate in the elaboration of the active sites (Fig. 2a). Analyses using the DALI server (Holm & Rosenström, 2010) indicate that the closest structural homologues of BsrV are alanine racemases from *Bacillus anthracis* (Alr<sub>Bax</sub>; Couñago *et al.*, 2009; r.m.s.d. of 2.0 Å for 361 C $\alpha$  atoms) and *E. coli* (Alr<sub>Ec</sub>; Wu *et al.*, 2008; r.m.s.d. of 2.5 Å for 347 C $\alpha$  atoms). Since the crystallization of AlrV was unsuccessful, further comparative analyses of BsrV were performed with Alr<sub>Ec</sub> (*E*-value of  $3 \times 10^{-157}$  to AlrV; Supplementary Fig. S1a). Superimposition of the BsrV and Alr<sub>Ec</sub> structures readily reveals their extensive similarity; nonetheless, substantial differences are also apparent (Fig. 2b and Supplementary Figs. S2 and S3).

The entry channel to the catalytic site is approximately twofold wider in BsrV than in Alr<sub>Ec</sub> (roughly  $10 \times 14$  Å in BsrV compared with  $7 \times 9.5$  Å in Alr<sub>Ec</sub>; Fig. 2b and Supplementary Figs. S2b and S2c). This difference is a direct consequence of shorter loops in the side walls of the catalytic entry site in BsrV and of the different orientation of helix  $\alpha 10$  (Fig. 2b and Supplementary Figs. S2b and S2c). In addition, the catalytic entry site has a negative charge in BsrV but not in Alr<sub>Ec</sub> (Supplementary Fig. S2d), perhaps impeding the entry of acidic substrates into the catalytic site and thereby contributing to the selectivity of BsrV for neutral and basic amino acids. BsrV also contains a wider tunnel connecting the entry to the catalytic centre (the narrowest width of the BsrV tunnel is 9.0 Å compared with 6.3 Å in Alr<sub>Ec</sub>).

The structure of BsrV also differs from that of Alr<sub>Ec</sub> owing to the presence of an additional structural element formed

**Table 2**  
Data-collection and refinement statistics.

Data set	BsrV	$\Delta$ Cl-BsrV	Bsr <sub>Ah</sub>	Alr <sub>Ah</sub>
Space group	<i>C</i> 2	<i>C</i> 2	<i>P</i> 2 <sub>1</sub>	<i>C</i> 222 <sub>1</sub>
Unit-cell parameters				
<i>a</i> (Å)	96.62	95.99	56.67	110.80
<i>b</i> (Å)	51.09	50.99	81.85	134.74
<i>c</i> (Å)	76.73	76.46	78.28	192.15
$\alpha$ (°)	90	90	90	90
$\beta$ (°)	101.15	100.69	98.97	90
$\gamma$ (°)	90	90	90	90
X-ray source	ID29, ESRF	X06SA, SLS	ID14-4, ESRF	ID29, ESRF
Data processing				
Temperature (K)	100	100	100	100
Wavelength (Å)	0.95994	0.99987	0.93927	0.97901
Resolution (Å)	47.3–1.15 (1.21–1.15)	44.8–1.50 (1.58–1.50)	42.1–1.45 (1.53–1.45)	48.04–3.25 (3.43–3.25)
Measured reflections	483260	428625	1150270	75325
Unique reflections	128145	57665	124710	21300
Multiplicity	3.3 (3.1)	6.7 (6.6)	7.4 (7.0)	3.5 (3.6)
Completeness (%)	98.5 (98.5)	99.0 (99.0)	99.9 (99.7)	93.3 (95.4)
Average <i>I</i> / $\sigma$ ( <i>I</i> )	11.5 (4.3)	14.7 (4.2)	12.8 (2.5)	11.5 (1.8)
Molecules in asymmetric unit	1	1	2	4
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	2.2	2.2	1.6	2.3
Solvent content (%)	43.6	43.6	25.0	47.8
<i>R</i> <sub>merge</sub>	0.05 (0.22)	0.08 (0.47)	0.09 (0.79)	0.09 (0.68)
Refinement				
Resolution range (Å)	47.3–1.15	47.1–1.50	42.1–1.45	47.32–3.25
Total No. of atoms	6670	6396	13365	10220
No. of protein non-H atoms	3074	2976	6238	9900
No. of ions (Cl <sup>-</sup> )	1	—	2	—
No. of water molecules	532	447	931	177
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.16/0.17	0.15/0.17	0.19/0.21	0.26/0.29
R.m.s.d., bond length (Å)	0.010	0.010	0.004	0.009
R.m.s.d., bond angles (°)	1.368	1.297	0.620	1.110
Ramachandran statistics, residues in				
Most favoured regions (%)	97.10	97.17	97.09	84.85
Allowed regions (%)	2.99	2.83	2.91	10.20
Disallowed regions (%)	0.25	0.0	0.0	4.95
PDB code	4beu	4beq	4bf5	4bhy

from the sequence at the N-terminus of the protein (residues 24<sup>4</sup>APLHIDT<sup>30</sup>). These amino acids fold into a stable, hinge-like complementary structure between monomers that shifts their relative positions (Figs. 2b and 2c). In BsrV monomers, the relative positions of the N-terminal domain (residues 24–279) and the C-terminal domain (residues 280–410) are offset by a 9° rotation compared with the equivalent domains in Alr<sub>Ec</sub> (Fig. 2c), and an average displacement of 1.5 Å for the whole protein backbone is observed. Consequently, there are dramatic differences between the disposition and conformation of the active sites of BsrV and those of Alr<sub>Ec</sub> (*e.g.* an increased distance between active sites from 37.6 Å in Alr to 45 Å in BsrV; Fig. 2c, right). The N-terminal structure of BsrV appears to be important for the activity of the enzyme, since a single mutation in this structure (P25E; Fig. 2b and Supplementary S3a) completely abolishes the activity of BsrV towards Ala and Met and dramatically impairs (70% reduction) its activity towards Arg.

### 3.4. The catalytic machinery of BsrV

In addition to the differences between the relative positions of protein domains in BsrV and Alr<sub>Ec</sub>, comparative structural analyses revealed significant differences between the catalytic

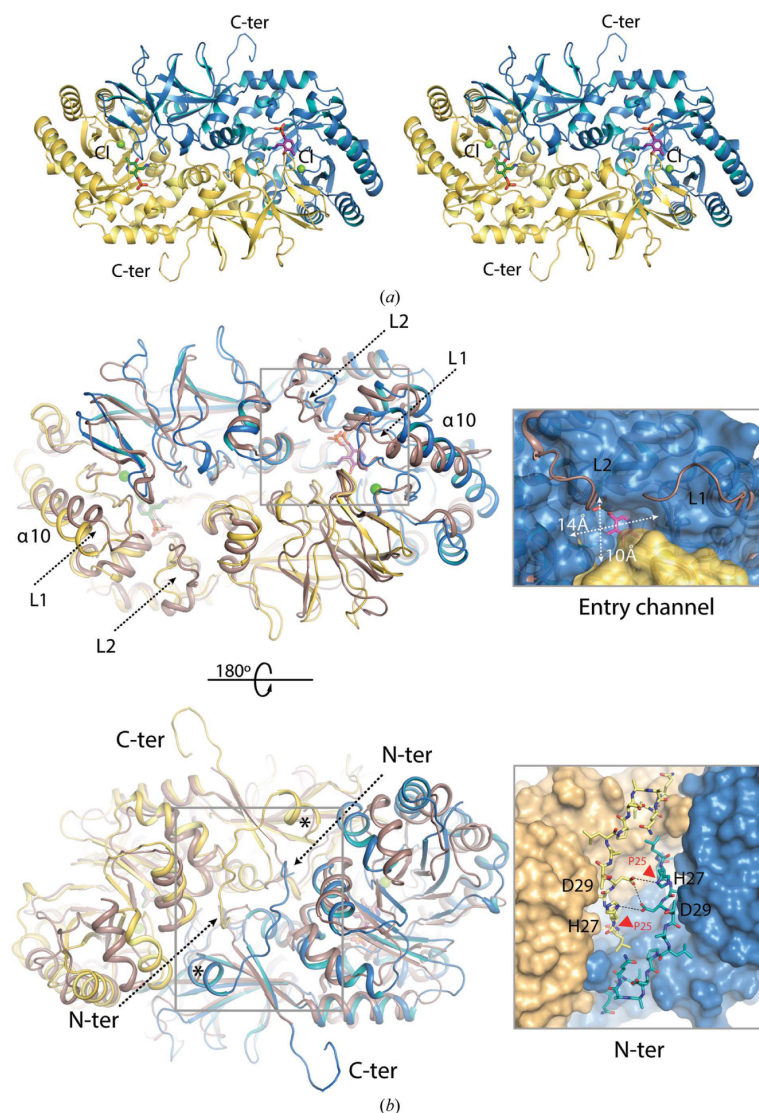
centres of the two enzymes. Notably, the enzymes use different means to position their PLP cofactor (Fig. 3*a* and Supplementary Fig. S3*b*). In BsrV, a Tyr residue that is conserved

among Alrs and is typically involved in coordination of the phosphate moiety of PLP is replaced by a proline residue (Pro391), thereby enlarging the space within the active site.

Phosphate coordination in BsrV instead depends on Ser245 and Tyr208 and a network of interactions mediated by water molecules (Fig. 3*a* and Supplementary Fig. S3*b*). No differences are observed in the stabilization of the pyridoxal ring, where Arg260 stabilizes the N atom of the ring through its N<sup>ε</sup> atom (2.82 Å).

The active site of BsrV also lacks an *N*-carboxylated lysine (Kcx) that in Alr<sub>Ec</sub> and most alanine racemases stabilizes an arginine involved in substrate binding within the active site (Watanabe *et al.*, 2002). In BsrV, Kcx is replaced by Ala165 and the substrate-binding Arg (Arg173) is then stabilized in the active site by the presence of an ion (Fig. 3*a* and Supplementary Figs. S3*b* and S3*c*). Its tetracoordination (with one water molecule at 3.28 Å, the N<sup>δ2</sup> atom of Asn174 at 3.58 Å, the N<sup>ε</sup> atom of Arg173 at 3.67 Å and the N<sup>η2</sup> atom of Arg173 at 3.61 Å) and low *B* factor suggested that it was chloride (Cl<sup>−</sup>; Fig. 3*a* and Supplementary S3*c*). A Cl<sup>−</sup> ion has also been found in the monospecific alanine racemase Alr<sub>Bax</sub>; however, unlike in BsrV, coordination of Cl<sup>−</sup> in Alr<sub>Bax</sub> is mediated by the same residues that interact with Kcx in Alr<sub>Ec</sub> and other enzymes that contain Kcx (Couñago *et al.*, 2009). Furthermore, the position occupied by Asn174 in BsrV (Fig. 3*a*) contains hydrophobic residues (Leu and Ile) in Alr<sub>Ec</sub>, Alr<sub>Bax</sub> and other monospecific Alrs. Collectively, these analyses provide strong evidence that the presence of a negative charge near the catalytic residue is critical for amino-acid racemization by these enzymes; however, they also illuminate the differing catalytic environments in the active sites of enzymes with distinct specificities.

The relevance of the Cl<sup>−</sup> to BsrV activity was further investigated by the structural and biochemical characterization of a BsrV R<sub>173</sub>N<sub>174</sub>/AA point mutant (ΔCl-BsrV). As expected, analysis of the crystal structure of ΔCl-BsrV (Table 2 and Fig. 3*b*) revealed that the Cl<sup>−</sup> was absent. However, the structure of ΔCl-BsrV



**Figure 2**

Three-dimensional structure of the BsrV dimer. (a) Crystal structure of a BsrV dimer with monomers depicted in yellow and blue. PLP molecules are shown in sticks and Cl<sup>−</sup> ions are shown as green spheres. (b) Structural superimposition of BsrV coloured as in (a) and Ala racemase from *E. coli* (Alr<sub>Ec</sub>) in brown. Loops L1 and L2 and  $\alpha$ -helix  $\alpha$ 10 at the entry channel are labelled [six residues in L1 (204–210) and 11 residues in L2 (267–278) in BsrV *versus* ten residues in L1 (159–169) and 18 residues (216–234) in L2 in Alr<sub>Ec</sub>]. On the right, differences in their active-site cavities are shown. White arrows show the dimensions of the entrance of the BsrV cavity. At the bottom of (b), a 180° rotation view (back side) of the dimer is shown and the BsrV N-terminal insertion is highlighted in the box (bottom right). The position of the N-terminal residue for both monomers in the Alr<sub>Ec</sub> structure is marked by an asterisk. In the close-up view of the back-side region of BsrV, the N-terminal insertion is shown as sticks. Polar interactions between the N-terminal extensions from both monomers are represented as dotted lines. The position of Pro25 is indicated by a red arrow.



## research papers

was otherwise very similar to that of BsrV; the only minor changes observed were in the side chains of Tyr208 and His204 (Fig. 3*b*). As anticipated,  $\Delta$ Cl-BsrV had a highly reduced or no ability to racemize most amino acids (Fig. 3*c* and Supplementary Figs. S3*d* and S3*e*), consistent with Cl<sup>−</sup> and Arg173 serving to replace Kcx in stabilizing a variety of substrates in the active site of BsrV.

### 3.5. Identification of molecular signatures that distinguish BsrV-like and Alr-like racemases

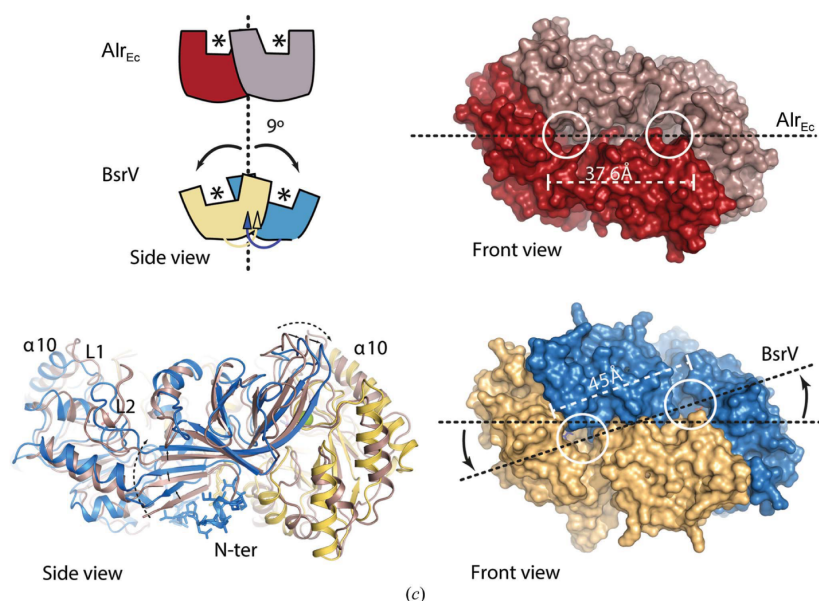
We hypothesized that the basis for the broad substrate tolerance of BsrV might be recognizable through sequence analysis and therefore searched for a molecular footprint of residues that are conserved among BsrV-like enzymes and differ from those found in monospecific alanine racemases. To initiate this analysis, a representative, nonredundant, set of PLP-dependent amino-acid racemase sequences (Alr and Bsr) from different bacteria were collected and aligned. Multiple sequence alignment (MSA; see §2) revealed differential conservation patterns that allowed clustering of proteins into three subfamilies (Fig. 4*a*). Subfamily 1 (in red) is composed of 84 proteins and includes AlrV, Alr<sub>Ec</sub> and Alr2 (DadB) from *E. coli*, which have been experimentally confirmed as alanine-specific racemases (Fig. 1*c*; Faraci & Walsh, 1988; Lam *et al.*, 2009). Subfamily 2 (in blue) is comprised of 13 proteins,

including BsrV (Alr2 from *V. cholerae*). Of the remaining 39 proteins in the data set, 24 proteins were grouped as subfamily 3 (in green). The 15 outlier proteins (in black) could not be assigned to any of the subfamilies (Fig. 4*a*). Based on the grouping of the characterized racemases, we assumed that the proteins in subfamily 1 are monospecific alanine racemases, while the proteins in subfamily 2 have a broader substrate spectrum. As no member of subfamily 3 or any of the 15 outlier proteins has been biochemically characterized, these proteins were not considered for further sequence analysis.

Further comparative analyses of subfamilies 1 and 2 (see Supplementary Material) enabled the prediction of an initial set of 31 specificity-determining positions (SDPs) that appeared to typify the specific and broad-spectrum racemase subfamilies. After additional refinement with manual inspection of each potential SDP, this approach allowed the selection of 16 sites within these enzymes that appear to constitute a molecular footprint that can be used to distinguish alanine and broad-spectrum racemases (Fig. 4*b*; discussed further below). Mapping of the broad-spectrum-linked amino-acid loci to the BsrV structure revealed that these amino acids affect or reside within features that distinguish BsrV from AlrV, including placement of the entry sites (Arg119 and Arg121), the size of the catalytic entry site (Pro206, Tyr208, Lys216 and Tyr246) and the catalytic machinery (Cys70, Ala165, Asn167, Gly169, Asn174, Gly263, Asn348, Thr349 and Pro391; see Fig. 5*h*).

Notably, specificity determinants were not found within the N-terminal extension of BsrV, although all of the putative broad-spectrum enzymes appear to possess this signature in the mature protein (relative to Alrs) as well as a putative signal sequence (Supplementary Table S3).

An expanded dataset containing 2540 putative PLP racemases was similarly grouped into subfamilies and the identities of these families were determined based on the enrichment of proteins described as 'broad-spectrum' or 'specific' racemases in the previous analysis (see Supplementary Material). This strategy resulted in the identification of 74 BsrV-like racemases that had probably been misannotated previously as being specific racemases. Intriguingly, most Bsr candidates are found in Gram-negative marine bacteria, primarily (54.5%) in members of the Vibrionaceae family (Supplementary Table S4).



**Figure 2 (continued)**

(*c*) Different oligomeric arrangement between BsrV and Alr<sub>Ec</sub> and its effect on the spatial distribution of the active sites. A schematic representation of the Alr<sub>Ec</sub> and BsrV dimers is shown on the left; asterisks mark active sites. A rotation of 9° between monomers is produced by the N-terminal insertion in BsrV. Bottom left, structural superimposition of BsrV and Alr<sub>Ec</sub> (brown cartoon). Displacement of the BsrV structural core by the N-terminal insertion (sticks) is represented by black arrows. Molecular surfaces for Alr<sub>Ec</sub> and BsrV are shown on the right. The positions of active sites are marked by white circles and differences in distances and position of BsrV active sites *versus* Alr are highlighted. See also Supplementary Fig. S2.



### 3.6. The structure of the *A. hydrophila* broad-spectrum racemase mirrors that of BsrV

To validate our *in silico* identification of putative broad-spectrum racemases, we performed biochemical and structural characterization of *A. hydrophila* Alr3 (a putative Bsr,

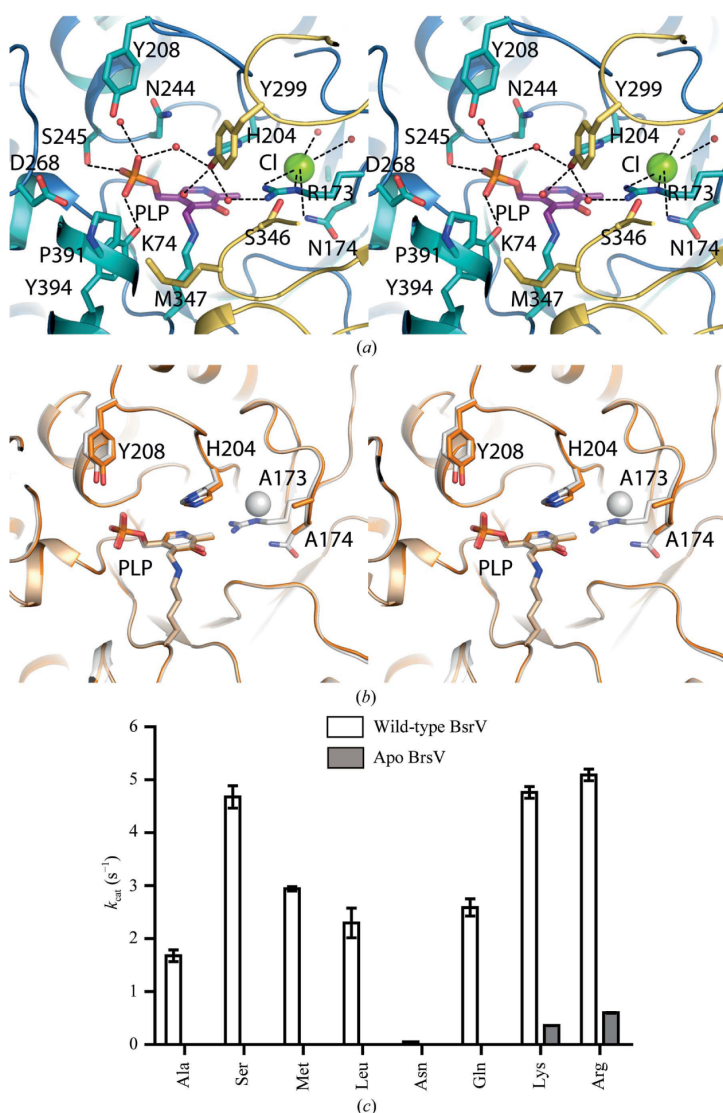
renamed Bsr<sub>Ah</sub>) and the putative primary alanine racemase of *A. hydrophila* (Alr<sub>Ah</sub>), which were identified from the 'broad-spectrum' and 'specific' subfamilies, respectively (Figs. 4a and 4c, Table 2 and Supplementary Fig. S4 and Table S4). As predicted, structural comparison revealed that Alr<sub>Ah</sub> and Bsr<sub>Ah</sub> differ in a manner analogous to that seen for BsrV and

Alr<sub>Ec</sub> (Supplementary Fig. S4), and Bsr<sub>Ah</sub> exhibits all of the distinguishing structural features of BsrV [*i.e.* Cl<sup>−</sup> instead of Kcx, a wider entry channel (shorter flanking loops) for the catalytic site, an N-terminal hinge-like extension and conserved signature residues; Fig. 4c and Supplementary Figs. S4d–S4h]. Bsr<sub>Ah</sub> also resembles BsrV in its ability to discriminate between amino-acid substrates and displayed catalytic activity against an identical subset of natural amino acids (Supplementary Figs. S4i and S4j). Therefore, the distinct subfamilies of racemases evident in bioinformatic analyses appear to comprise a functionally distinct subset of amino-acid racemases with broad substrate specificity.

### 3.7. Molecular modelling of the docking of diverse amino-acid substrates within the BsrV and Alr<sub>Bs</sub> active sites

To unveil the structural determinants that influence substrate stabilization in specific and broad-range racemases, we took advantage of the crystal structure of *B. stearotherophilus* alanine racemase (Alr<sub>Bs</sub>) in complex with *N*-(5'-phosphopyridoxyl)-D-alanine (PLP-D-Ala; Watanabe *et al.*, 2002; PDB entry 1lg6) as the reference since this structure mimics the enzyme–substrate catalytic intermediate. We superimposed Alr<sub>Bs</sub> in complex with PLP-D-Ala on BsrV and substituted the alanine analogue by all of the different kinds of amino-acid substrates, selecting the most stable conformation in each case (Figs. 5a–5f). Whilst Alr<sub>Bs</sub> could only accommodate alanine and serine (not shown) in its catalytic site (Figs. 5b, 5d and 5f and Supplementary Table S5b), the BsrV catalytic pocket appears to accommodate all ten amino acids found to be racemized by this enzyme (Figs. 5a, 5c and 5e). Furthermore, in full agreement with the substrate specificity of BsrV, the remaining nine amino acids present multiple steric clashes when forced to fit into the active site of the enzyme (Supplementary Fig. S5a).

The docking analysis supported two overlapping mechanisms for amino-acid



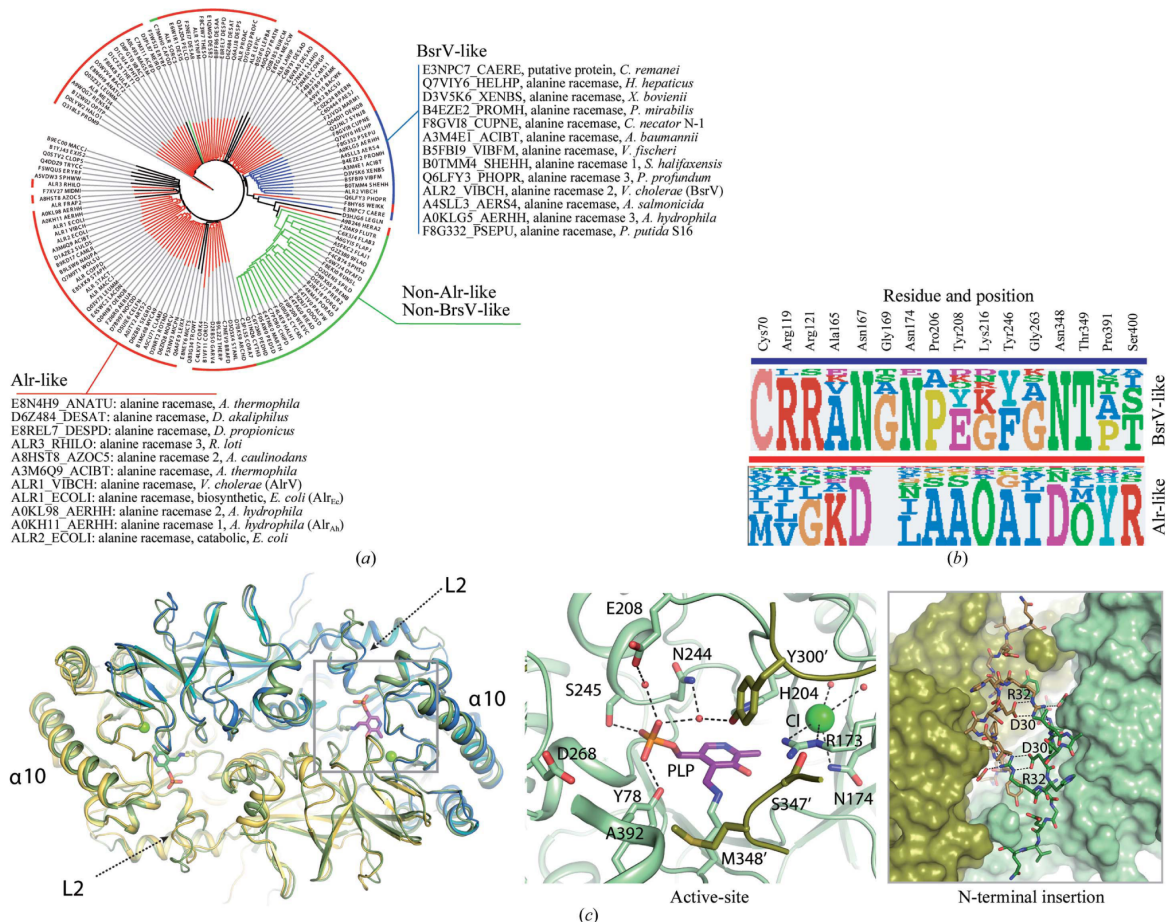
**Figure 3** BsrV and  $\Delta$ Cl-BsrV active-site structures and kinetic parameters. (a) Stereoview of the BsrV active site. Relevant residues and PLP are shown as sticks; water molecules and Cl<sup>−</sup> ions are shown as red and green spheres, respectively. (b) Stereoview showing the structural superimposition of the active sites of wild-type BsrV (shown as a white transparency) and  $\Delta$ Cl-BsrV (R<sub>173</sub>N<sub>174</sub>/AA) (coloured orange). No other significant changes apart from the double mutation and consequent Cl<sup>−</sup> loss (white sphere) are apparent. (c) Comparison between the  $k_{cat}$  of  $\Delta$ Cl-BsrV and BsrV for different racemizable amino acids. The results in (c) are means  $\pm$  SD of triplicates from two independent experiments. See also Supplementary Fig. S3.

## research papers

stabilization by BsrV: one for aliphatic residues and another for basic amino acids. Aliphatic amino acids (Ala, Ser, Cys, Leu, Gln and Met) are likely to be stabilized in the active site of BsrV through polar interactions between their carboxylate moiety and Arg173, Tyr299' and Tyr318', in a similar way to that defined for Ala in Alr<sub>Bs</sub> (Figs. 5a and 5b). In addition, the amino-acid side chains are stabilized through hydrophobic interactions with Met347' in BsrV (Figs. 5a, 5c and 5e). Basic amino acids (Arg, Lys and Orn) are stabilized by the same interactions at their carboxylate moieties and, in addition, by hydrogen bonding to Tyr394 and by electrostatic interactions with Asp268 and the PLP phosphate group (Fig. 5e).

Furthermore, docking results neatly explain why BsrV, but not Alr<sub>s</sub>, racemizes long aliphatic and basic residues. In Alr<sub>s</sub>, a Tyr residue (Tyr354 in Alr<sub>Bs</sub>) closes the active-site cavity, which accommodates the short chain of the Ala residue (Fig. 5b) but produces steric clashes when longer aliphatic, or basic, residues are docked (Figs. 5d and 5f). The presence of a Pro residue (Pro391) at the equivalent position in BsrV makes sufficient additional space for the proper fitting of both long aliphatic and basic residues (Fig. 5c and 5e).

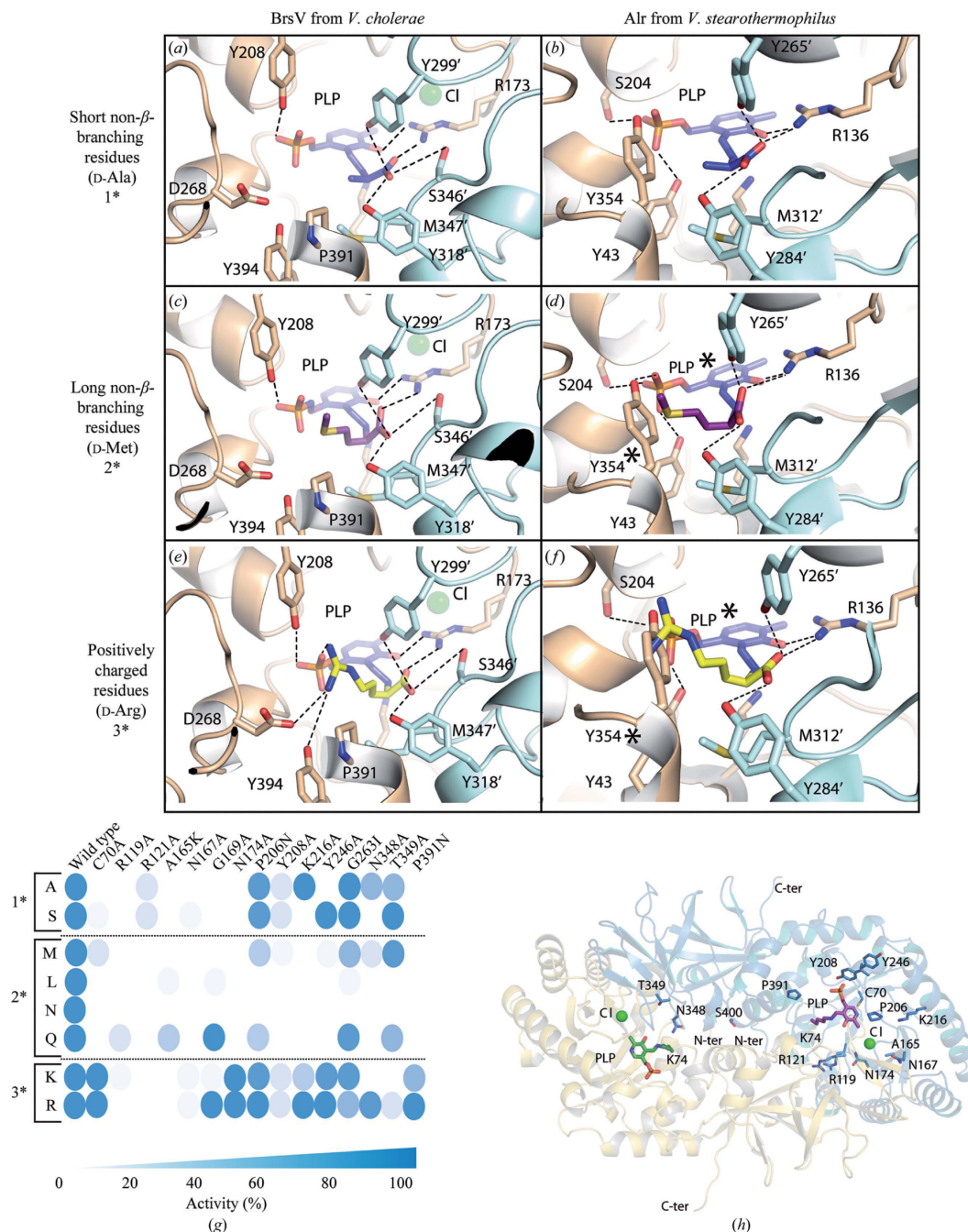
The differential efficiency of BsrV for its substrates was also in agreement with the docking results. Indeed, the lower activity values found for His and Asn (Supplementary Fig.



**Figure 4**

The identification of 16 residues that constitute the molecular footprint for BsrV multispecificity and experimental validation. (a) Tree of BsrV orthologues obtained through bioinformatic sequence analysis (see §2). Alr-like enzymes are highlighted in red, Bsr-like enzymes in blue and non-Alr-like, non-Bsr-like enzymes in green; 15 outlier proteins that do not group with any of these families are shown in black. All Bsr-like family members (including BsrV) and a small subset of Alr-like enzymes (including AlrV and Alr<sub>ec</sub>) are listed. (b) Molecular footprint. Sequence composition (displayed as sequence logos) of the specificity-determining positions for Alr/BsrV-like racemases. The specific residues and their positions are listed at the top. (c) Structural analysis of the broad-spectrum racemase from *A. hydrophila* (Bsr<sub>Ah</sub>). A structural superimposition of BsrV with Bsr<sub>Ah</sub> is shown on the left. Loop L2 and the  $\alpha$ -helix  $\alpha$ 10 at the entry channel are labelled. The central panel shows a close-up view of the Bsr<sub>Ah</sub> catalytic site. Relevant residues and PLP are shown as sticks; water molecules and Cl<sup>-</sup> ions are shown as red and green spheres, respectively. In the right box, a close-up view of the back-side region (180° rotation view) of Bsr<sub>Ah</sub> with the N-terminal insertion in stick representation is shown. Polar interactions between relevant residues of the N-terminal extensions from both monomers are represented as dotted lines. See also Supplementary Fig. S4.





**Figure 5**

DAA recognition by BsrV from *V. cholerae* versus Alr from *B. stearothermophilus*. Docked models of enzyme–substrate catalytic intermediates (BsrV and Alr<sub>FB</sub>) for (a, b) short non- $\beta$ -branching residues (D-Ala, shown in stick representation with C atoms coloured blue) (1\*), (c, d) long non- $\beta$ -branching residues (D-Met, shown in stick representation with C atoms coloured magenta) (2\*) and (e, f) positively charged residues (D-Arg, shown in stick representation with C atoms coloured yellow) (3\*). Protein residues involved in substrate recognition are labelled and interactions are highlighted as dotted lines. Substrates clashes are labelled by an asterisk. (g) Schematic of specific activity (relative to wild-type BsrV; left column) of BsrV single mutants in the ‘signature’ SDP. The results correspond to means of triplicates. (h) Location of the SDP within the BsrV structure. See also Supplementary Fig. S5.

## research papers

S1g) can be explained by the limited interactions with Asp268 owing to their shorter side chains (Supplementary Figs. S5c and S5d). Nonetheless, additional elements seem to participate in the stabilization of basic substrates.

### 3.8. Functional verification of the putative specificity determinants of BsrV

To further understand the different specificities of BsrV and AlrV and to assess the significance of each 'signature' residue, we examined the activity of purified BsrV containing individual point mutations at these loci (Figs. 5g and 5h and Supplementary Figs. S5f and S5g). Notably, the mutation of any of the signature sites reduced the activity of BsrV, although the extent of the reduction varied among the mutations and among the classes of substrates (*i.e.* short and long non- $\beta$ -branching aliphatic chains *versus* positively charged side chains; Fig. 5g and Supplementary S5g). Structural analysis provides likely explanations for the effects of these mutations on BsrV activity (Fig. 5h and Table 3). The analysis of the activity of each BsrV mutant isoform revealed that five substitutions (R119A, R121A, A165K, N167A and Y208A) reduced the activity below 20% compared with wild-type BsrV for most of the substrates (Fig. 5g). The footprint residues Arg119 and Arg121 are involved in salt-bridge interactions on the back side of the dimer and allow the formation of the N-terminal extension of BsrV. The dramatic effect of their absence confirms our previous conclusion (based on the P25E mutant) that the N-terminal extension is critical for the activity of BsrV towards a variety of substrates. In contrast, the A165K and N167A substitutions are expected to impair  $\text{Cl}^-$  coordination and consequently the stabilization of the catalytic Arg173. Finally, the Y208A mutant should affect the conformation of the entry loop L1 (Table 3) and, as expected, has an especially large effect on the racemization of amino acids with long side chains.

The remaining substitutions have narrower effects, and most frequently limit the activity of BsrV towards long non- $\beta$ -branching amino acids, possibly because these substrates encounter more steric obstacles in entering and/or being positioned within the active site. Interestingly, a subset of mutations (G169A, N174A and P391N) that prevent the racemization of non- $\beta$ -branching aliphatic amino acids do not markedly impair the activity of BsrV towards basic substrates. Based on the docking models, it seems likely that the side chains of basic substrates are stabilized by different components within the catalytic site (*e.g.* PLP, Asp268 and Tyr394; Fig. 5e). Analysis of a D268N mutant further confirmed that Asp268 contributes to the racemization of basic amino acids by BsrV (Supplementary Fig. S5e).

**Table 3**  
Structural analysis of the molecular footprint of BsrV.

Mutation	Protein region	Structural explanation
C70A	Active site	Structural stabilization of active site by a disulfide bridge
R119A	N-terminal insertion	Arg119 is involved in salt-bridge interactions allowing N-terminal insertion in BsrV
R121A	N-terminal insertion	Arg121 is involved in salt-bridge interactions allowing N-terminal insertion in BsrV
A165K	Active site	Ala at this position allows the correct stabilization of the $\text{Cl}^-$ ion by Arg173 and Asn174 (in Alr, this position is occupied by the N-carboxylated Lys)
N167A	Active site	Structural stabilization of loop involved in $\text{Cl}^-$ coordination
G169A	Active site	Loop involved in $\text{Cl}^-$ coordination
N174A	Active site	Residue involved in $\text{Cl}^-$ coordination
P206N	Entry loops	Structural stabilization of the L1 loop (involved in substrate entry)
Y208A	Entry loops	Stabilization of the PLP phosphate group
K216A	Entry loops	Salt-bridge interaction stabilizing $\alpha 10$ at the entry
Y246A	Entry loops	Mutation will block movement of the r2 region
G263I	Active site	Gly residue allows Asp268 and Tyr394 stabilization of side chains of different substrates
N348A	Active site	Stabilization of loop that includes Met347 at the catalytic site
T349A	Active site	
P391N	Active site	This Pro is critical in the broad specificity of BsrV, allowing the stabilization of side chains from different substrates

### 4. Discussion

Broad-spectrum amino-acid racemases such as BsrV enable bacteria to generate NCDAA, the varied roles of which in microbial physiology are just beginning to be appreciated (Cava, de Pedro *et al.*, 2011; Kolodkin-Gal *et al.*, 2010, 2012; Lam *et al.*, 2009). Here, we present extensive crystallographic, biochemical and bioinformatic analyses of two broad-spectrum racemases (BsrV and Bsr<sub>AB</sub>) and identify the structural features that enable such enzymes to accommodate a wider range of substrates than related PLP-dependent alanine racemases. We also identify the molecular signature for BsrV-like racemases from diverse microbes that enables them to be distinguished from monospecific alanine racemases.

Although BsrV from *V. cholerae* has previously been reported to produce primarily D-Met and D-Leu (Lam *et al.*, 2009), we found that this enzyme can also racemize eight additional amino acids typically found in proteins, as well as several nonproteinogenic amino acids. The substrates of BsrV fall into two main classes: non- $\beta$ -branched aliphatic amino acids (*i.e.* Ala, Ser, Leu, Cys, Gln, Asn and Met) and basic amino acids (*i.e.* Arg, Lys and His). Although D forms of the latter class were not detected *in vivo*, kinetic analyses indicate that they can be readily generated by BsrV, perhaps facilitated by additional stabilizing interactions between these substrates and the active site of BsrV (Figs. 5a, 5c and 5e). Lys and Arg have also been found to be the preferred substrates of a broad-spectrum periplasmic racemase from *Pseudomonas taetrolens* (Matsui *et al.*, 2009). The enhanced active-site stabilization predicted for basic substrates may account for the resilience of the catalytic activity of mutant BsrV with such substrates relative to large uncharged aliphatic substrates (Fig. 5g). The substrate range of BsrV, which includes activity towards non-natural substrates (*e.g.* ornithine, norleucine, homoserine, N-acetyl lysine methyl ester, diaminobutyrate and aminobutyrate; Supplementary Fig. S1e), is broader than

any other known amino-acid racemase and suggests that it has great potential for biotechnological and industrial applications. Currently, production of DAA is an expensive process that is typically reliant upon inefficient chemical catalysts (Gong *et al.*, 2010; Green *et al.*, 2009; Toth & Richard, 2007).

Interestingly, all of the Bsr-family racemases are encoded by Gram-negative bacteria. Since some Gram-positive organisms (*e.g.* *B. subtilis*) have also been reported to produce NCDAA, this result suggests that additional families of broad-spectrum enzymes may remain to be identified. Consistent with this idea, the spectrum of NCDAA produced by *B. subtilis* includes several amino acids that cannot be generated by BsrV (*e.g.* D-Tyr and D-Trp; Kolodkin-Gal *et al.*, 2010). Additional broad-spectrum enzymes may not be related to the PLP-dependent racemases such as Alrs and BsrS, but might instead bear a resemblance to other known racemases, such as MurI.

Comparative structural analyses of BsrV-family broad-spectrum racemases (BsrV and Bsr<sub>Ah</sub>) and more restricted enzymes (Alr<sub>Ah</sub>, Alr<sub>Ec</sub> and Alr<sub>Bs</sub>) suggest that the catalytic plasticity of these racemases has three essential determinants: the arrangement and features of the entry site, the width of the channel leading to the active site and its catalytic environment, which is able to accommodate and stabilize substrates of diverse chemical natures. There is no indication that the underlying mechanism of racemization differs between the two enzyme subfamilies since the dominant catalytic moieties (catalytic Lys and Tyr) are found in both. The wider entry site and channel of BsrV are likely to facilitate interaction with amino-acid substrates larger than alanine; indeed, structural, biochemical and modelling analyses suggest that the entry channel of BsrV can also accommodate peptide chains which are markedly larger. The catalytic site of BsrV-like racemases is also less constrained than those of related alanine racemases, apparently owing both to differences in the components of the chambers (*e.g.* Cl<sup>−</sup> and Pro391) and to different arrangements of the enzymatic domains that interact to form them. Previous analyses have revealed variability in the relative positioning of racemase N- and C-terminal domains owing to changes in the amino acids at the dimer interface (Couñago *et al.*, 2009). In broad-spectrum racemases, this interface is markedly altered by the presence of a new structure formed from an extended N-terminal sequence. Mutational analyses indicate that the N-terminal extension, as well as the 16 'signature' amino acids that can be used to identify broad-spectrum enzymes, are critical for racemase activity on the full set of substrates. Interestingly, mutations in most signature loci did not revert BsrV to an alanine-specific racemase; instead, its activity towards large aliphatic side chains was often lost, while its activity towards basic amino acids was preserved.

Collectively, given the impact of NCDAA on a variety of cellular processes (*e.g.* biofilm stability, sporulation and cell communication (Anfora *et al.*, 2007; Kolodkin-Gal *et al.*, 2010; O'Connor & Zusman, 1997), structural and regulatory characterization of broad-spectrum racemases might open the door to the design of new drugs (Conti *et al.*, 2011) that would challenge the adaptability of pathogenic bacteria to adverse conditions (*i.e.* during infection).

We thank Stavroula Hatzios for insightful comments on the manuscript and Lourdes Infantes for help in data-mining analysis. Special thanks to F. Lopez-Gallego for providing DAAO. Research in the Cava laboratory is supported by the MINECO, Spain (RYC-2010-06241), Universidad Autónoma de Madrid (UAM-38) and by the Knut and Alice Wallenberg Foundation (KAW). Additionally, this work was supported by the BFU2011-25326 MEC grant (JAH), by the S2010/BMD-2457 grant from CAM (JAH) and by HHMI (MKW).

## References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
- Anfora, A. T., Haugen, B. J., Roesch, P., Redford, P. & Welch, R. A. (2007). *Infect. Immun.* **75**, 5298–5304.
- Austin, B. (2010). *Vet. Microbiol.* **140**, 310–317.
- Babtie, A., Tokuriki, N. & Hoffelder, F. (2010). *Curr. Opin. Chem. Biol.* **14**, 200–207.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Smart, O. S., Vonrhein, C. & Womack, T. O. (2011). *BUSTER*. Cambridge: Global Phasing Ltd.
- Cava, F., de Pedro, M. A., Lam, H., Davis, B. M. & Waldor, M. K. (2011). *EMBO J.* **30**, 3442–3453.
- Cava, F., Lam, H., de Pedro, M. A. & Waldor, M. K. (2011). *Cell. Mol. Life Sci.* **68**, 817–831.
- Chang, C. Y., Thompson, H., Rodman, N., Bylander, J. & Thomas, J. (1997). *Ann. Clin. Lab. Sci.* **27**, 254–259.
- Chen, C. (1996). *Oral Microbiol. Immunol.* **11**, 425–427.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chen, Y., Zhang, W., Shi, Q., Heseck, D., Lee, M., Mobashery, S. & Shoichet, B. K. (2009). *J. Am. Chem. Soc.* **131**, 14345–14354.
- Conti, P., Tamborini, L., Pinto, A., Blondel, A., Minoprio, P., Mozzarelli, A. & De Micheli, C. (2011). *Chem. Rev.* **111**, 6919–6946.
- Couñago, R. M., Davlieva, M., Strych, U., Hill, R. E. & Krause, K. L. (2009). *BMC Struct. Biol.* **9**, 53.
- Levie, R. de (2001). *How to Use Excel in Analytical Chemistry and in General Scientific Data Analysis*. Cambridge University Press.
- Delong, E. F., Franks, D. G. & Yayanos, A. A. (1997). *Appl. Environ. Microbiol.* **63**, 2105–2108.
- Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
- El-Hajj, Z. W., Allcock, D., Tryfona, T., Lauro, F. M., Sawyer, L., Bartlett, D. H. & Ferguson, G. P. (2010). *Ann. N. Y. Acad. Sci.* **1189**, 143–148.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Faraci, W. S. & Walsh, C. T. (1988). *Biochemistry*, **27**, 3267–3276.
- Gong, L., Mulcahy, S. P., Devarajan, D., Harms, K., Frenking, G. & Meggers, E. (2010). *Inorg. Chem.* **49**, 7692–7699.
- Green, B. A., Yu, R. J. & Van Scott, E. J. (2009). *Clin. Dermatol.* **27**, 495–501.
- Hanahan, D. (1985). *DNA Cloning: A Practical Approach*, Vol. 1, edited by D. M. Glover, pp. 109–135. Oxford: IRL Press.
- Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T. & Calderwood, S. B. (2012). *Lancet*, **379**, 2466–2476.
- Heidelberg, J. F. *et al.* (2000). *Nature (London)*, **406**, 477–483.
- Herbert, E. E. & Goodrich-Blair, H. (2007). *Nature Rev. Microbiol.* **5**, 634–646.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Horcajo, P., de Pedro, M. A. & Cava, F. (2012). *Microb. Drug Resist.* **18**, 306–313.

## research papers

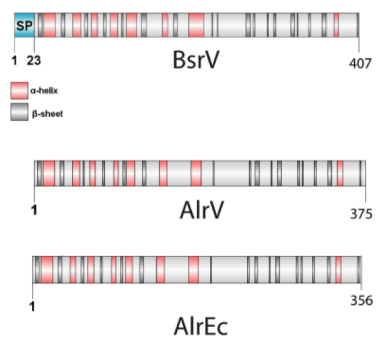
- Howard, A., O'Donoghue, M., Feeney, A. & Sleator, R. D. (2012). *Virulence*, **3**, 243–250.
- Inoue, H., Nojima, H. & Okayama, H. (1990). *Gene*, **96**, 23–28.
- Jacobsen, S. M. & Shirliff, M. E. (2011). *Virulence*, **2**, 460–465.
- Kaasch, A. J., Dinter, J., Goeser, T., Plum, G. & Seifert, H. (2012). *Infection*, **40**, 185–190.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Khersonsky, O., Roodveldt, C. & Tawfik, D. S. (2006). *Curr. Opin. Chem. Biol.* **10**, 498–508.
- Kim, S. E., Park, S. H., Park, H.-B., Park, K.-H., Kim, S.-H., Jung, S.-I., Shin, J.-H., Jang, H.-C. & Kang, S. J. (2012). *Chonnam Med. J.* **48**, 91–95.
- Kino, K., Sato, M., Yoneyama, M. & Kirimura, K. (2007). *Appl. Microbiol. Biotechnol.* **73**, 1299–1305.
- Kolodkin-Gal, I., Cao, S., Chai, L., Böttcher, T., Kolter, R., Clardy, J. & Losick, R. (2012). *Cell*, **149**, 684–692.
- Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R. & Losick, R. (2010). *Science*, **328**, 627–629.
- Komarova, N. V., Golubev, I. V., Khoronenkova, S. V., Chubar', T. A. & Tishkov, V. I. (2012). *Biochemistry*, **77**, 1181–1189.
- Laemmli, U. K. & Favre, M. (1973). *J. Mol. Biol.* **80**, 575–599.
- Lam, H., Oh, D.-C., Cava, F., Takacs, C. N., Clardy, J., de Pedro, M. A. & Waldor, M. K. (2009). *Science*, **325**, 1552–1555.
- LeMagueres, P., Im, H., Dvorak, A., Strych, U., Benedik, M. & Krause, K. L. (2003). *Biochemistry*, **42**, 14752–14761.
- LeMagueres, P., Im, H., Ebalunode, J., Strych, U., Benedik, M. J., Briggs, J. M., Kohn, H. & Krause, K. L. (2005). *Biochemistry*, **44**, 1471–1481.
- Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., Reinhardt, R., Brinkhoff, T., Harder, J., Amann, R. & Musmann, M. (2012). *ISME J.* **6**, 2178–2187.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125–132.
- Matsui, D. & Oikawa, T. (2010). *Chem. Biodivers.* **7**, 1591–1602.
- Matsui, D., Oikawa, T., Arakawa, N., Osumi, S., Lausberg, F., Stäbler, N., Freudl, R. & Eggeling, L. (2009). *Appl. Microbiol. Biotechnol.* **83**, 1045–1054.
- Murakami, K. (2012). *J. Gastroenterol.* **47**, 724–725.
- Nelson, M. D. & Fitch, D. H. (2011). *Methods Mol. Biol.* **772**, 459–470.
- Noda, M., Matoba, Y., Kumagai, T. & Sugiyama, M. (2004). *J. Biol. Chem.* **279**, 46153–46161.
- O'Connor, K. A. & Zusman, D. R. (1997). *Mol. Microbiol.* **24**, 839–850.
- O'Hara, C. M., Brenner, F. W. & Miller, J. M. (2000). *Clin. Microbiol. Rev.* **13**, 534–546.
- Pepperell, C., Kus, J. V., Gardam, M. A., Humar, A. & Burrows, L. L. (2002). *Antimicrob. Agents Chemother.* **46**, 3555–3560.
- Pinhassi, J., Pujalte, M. J., Macián, M. C., Lekunberri, I., González, J. M., Pedrós-Alió, C. & Arahal, D. R. (2007). *Int. J. Syst. Evol. Microbiol.* **57**, 2370–2375.
- Richard, J. P., Amyes, T. L., Crugeiras, J. & Rios, A. (2009). *Curr. Opin. Chem. Biol.* **13**, 475–483.
- Rømer Villumsen, K., Dalsgaard, I., Holten-Andersen, L. & Raida, M. K. (2012). *PLoS One*, **7**, e46733.
- Rosenberg, A. H., Lade, B. N., Chui, D.-S., Lin, S.-W., Dunn, J. J. & Studier, F. W. (1987). *Gene*, **56**, 125–135.
- Różalski, A., Kwil, I., Torzewska, A., Baranowska, M. & Staccek, P. (2007). *Postepy Hig. Med. Dosw. (Online)*, **61**, 204–219.
- Seifert, H., Strate, A., Schulze, A. & Pulverer, G. (1993). *Clin. Infect. Dis.* **17**, 632–636.
- Shaw, J. P., Petsko, G. A. & Ringe, D. (1997). *Biochemistry*, **36**, 1329–1342.
- Strych, U. & Benedik, M. J. (2002). *J. Bacteriol.* **184**, 4321–4325.
- Su, Y.-C. & Liu, C. (2007). *Food Microbiol.* **24**, 549–558.
- Sugar, D. R., Murfin, K. E., Chaston, J. M., Andersen, A. W., Richards, G. R., deLéon, L., Baum, J. A., Clinton, W. P., Forst, S., Goldman, B. S., Krasomil-Osterfeld, K. C., Slater, S., Stock, S. P. & Goodrich-Blair, H. (2012). *Environ. Microbiol.* **14**, 924–939.
- Timoney, P. J. (1996). *Comp. Immunol. Microbiol. Infect. Dis.* **19**, 199–204.
- Toth, K. & Richard, J. P. (2007). *J. Am. Chem. Soc.* **129**, 3013–3021.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Vollmer, W., Blanot, D. & de Pedro, M. A. (2008). *FEMS Microbiol. Rev.* **32**, 149–167.
- Watanabe, A., Yoshimura, T., Mikami, B., Hayashi, H., Kagamiyama, H. & Esaki, N. (2002). *J. Biol. Chem.* **277**, 19166–19172.
- Williamson, E. D. & Oyston, P. C. (2012). *J. Med. Microbiol.* **61**, 911–918.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.
- Winn, M. D. et al. (2011). *Acta Cryst.* **D67**, 235–242.
- Wu, D., Hu, T., Zhang, L., Chen, J., Du, J., Ding, J., Jiang, H. & Shen, X. (2008). *Protein Sci.* **17**, 1066–1076.
- Zhao, J.-S., Manno, D., Leggiadro, C., O'Neil, D. & Hawari, J. (2006). *Int. J. Syst. Evol. Microbiol.* **56**, 205–212.

## **Supplementary Material**

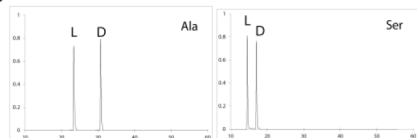
### **Structural bases for the broad specificity of a new family of amino acid racemases**

**Akbar Espaillat, Cesar Carrasco-López, Noelia Bernardo-Garcia, Natalia  
Pietrosemoli, Lisandro H. Otero, Laura Alvarez, Miguel A. de Pedro, Florencio  
Pazos, Brigid M. Davis, Matthew K. Waldor, Juan A. Hermoso and Felipe Cava**

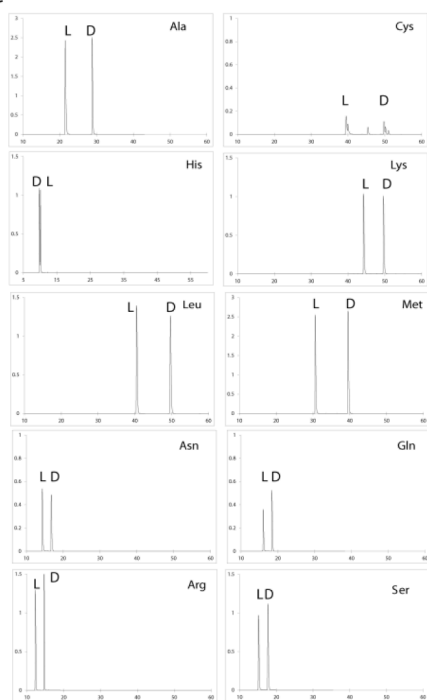
A



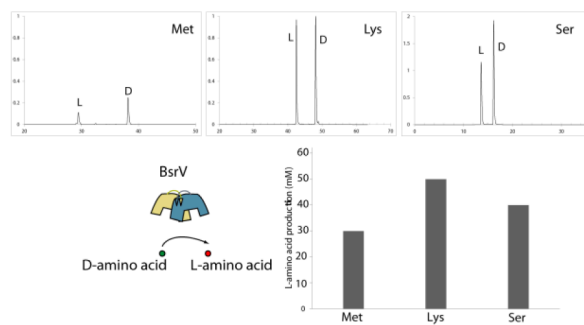
B



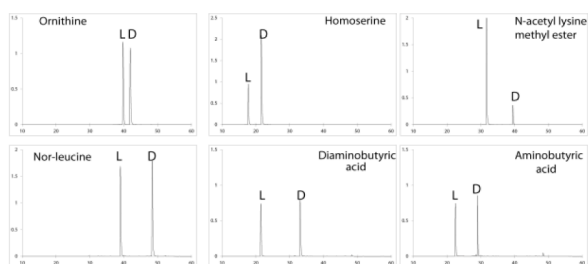
C



D



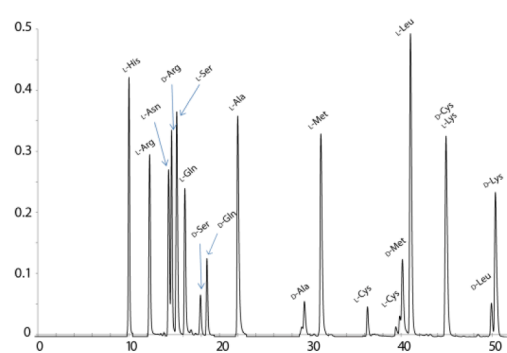
E



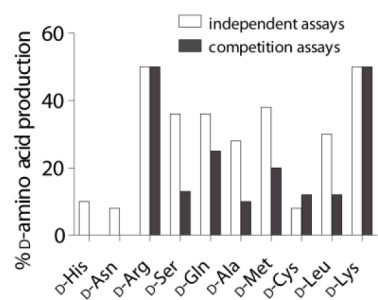
F

BsrV (no His-tag)	$K_m$ (mM)	$k_{cat}$ ( $s^{-1}$ )	$k_{cat}/K_m$ ( $mM^{-1} s^{-1}$ ). $10^3$
Met	12±4	3.1±0.42	25.8
Arg	20±6	4.77±0.75	23.9

G

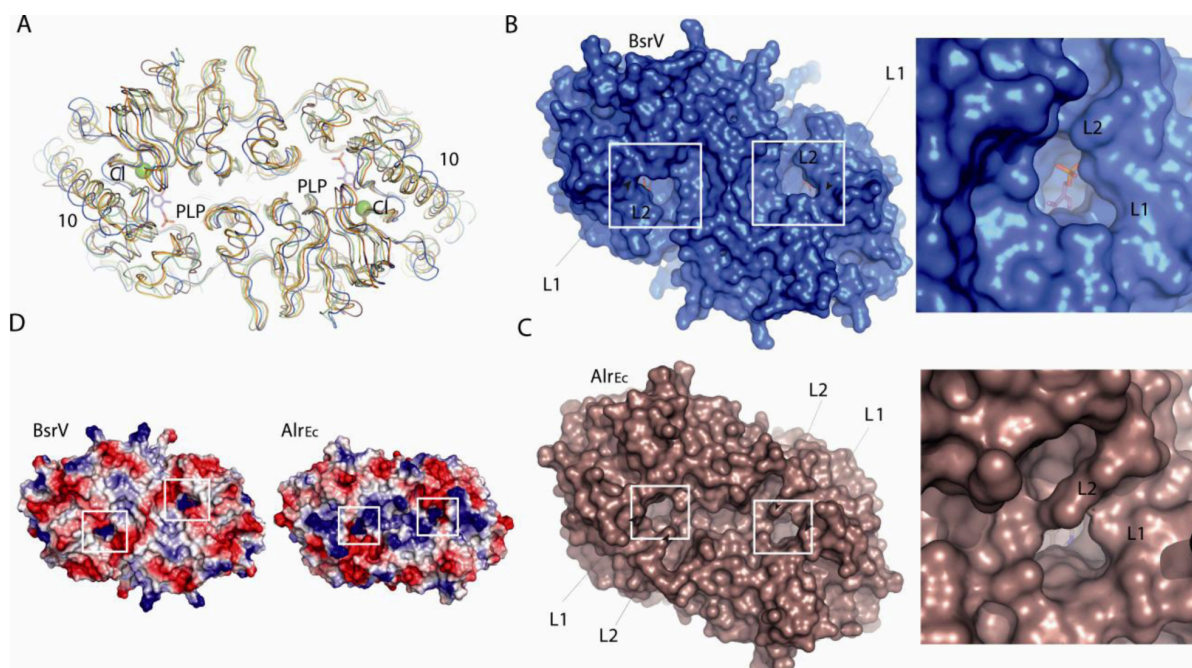


H



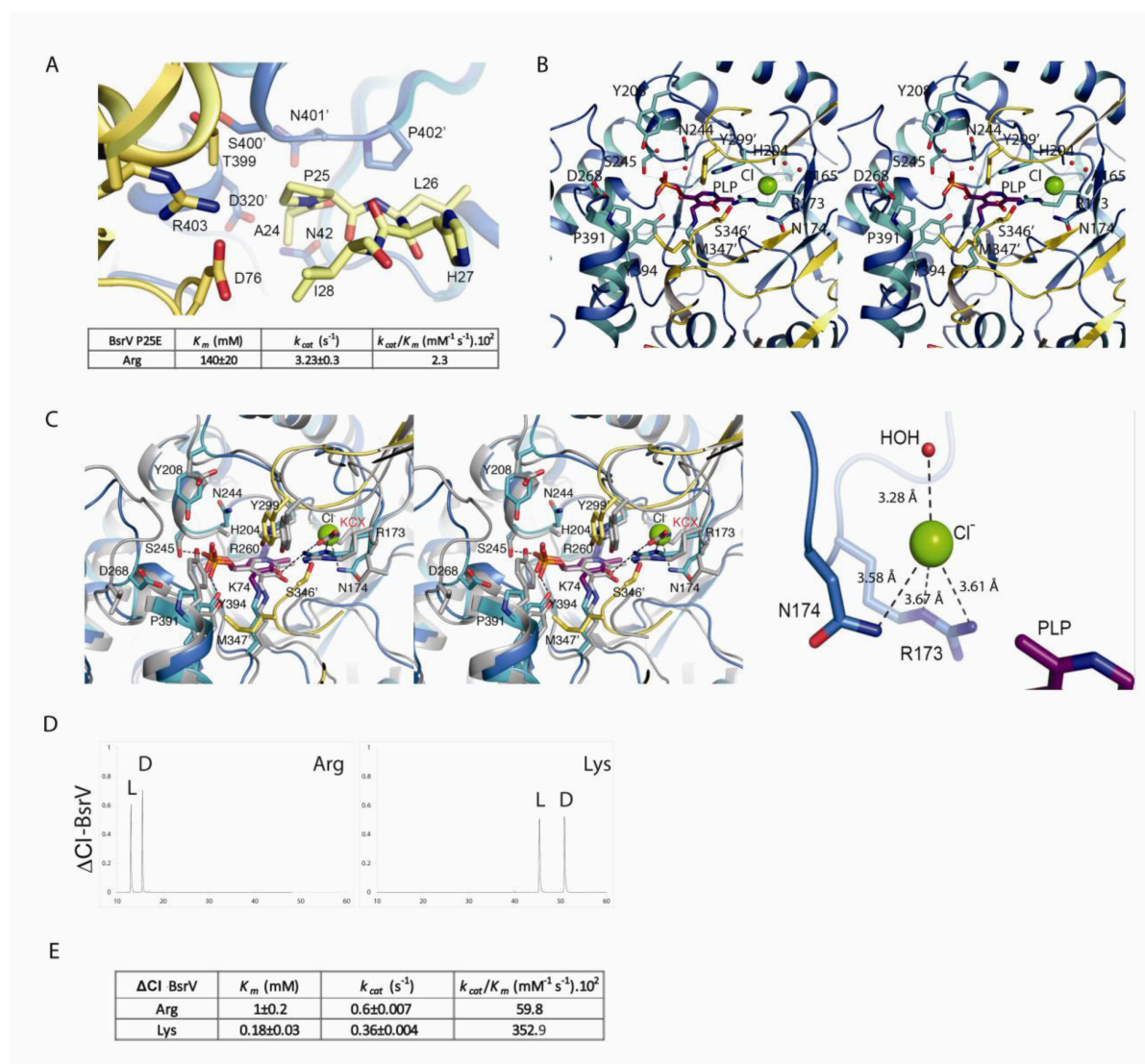


**Supplementary Figure S1.** Biochemical comparison of AlrV and BsrV from *V. cholerae*. (A) Prediction of protein domains for BsrV, AlrV and Alr<sub>Ec</sub>. Representation of protein features obtained with UNIPROT. In pink is shown the  $\alpha$ -helix, in grey the  $\beta$ -sheet and in blue the signal peptide (predicted with the online tool PSIPRED). (B-E) Racemization of amino acids by *V. cholerae*'s AlrV and BsrV. Chromatograms show HPLC analysis of Marfey's derivatized amino acids after *in vitro* racemase reactions; (B) AlrV and (C) BsrV activity on natural (proteinogenic) amino acids. (D) Reverse reaction starting from D-amino acids in BsrV. (E) BsrV activity on non-proteinogenic amino acids. (F) Kinetic values with Met and Arg of mature BsrV without His-tag. Note that kinetic analyses for representative amino acids yielded the same result using tagged (Fig. 1) and untagged protein. (G) Chromatograms of the derivatized products of a 5 minutes reaction containing BsrV and all natural racemizable amino acids. (H) Yield of DAA produced by BsrV in independent (white bars) and competition reactions (black bars).



**Supplementary Figure S2.** Structural comparison of multispecific BsrV and monospecific racemases. (A) Ribbon representation of BsrV (blue) superimposed on Ser racemase from *E. faecalis* (PDB code 4ECL) colored in green, Ala racemase from *B. anthracis* (PDB code 2VD8) colored in yellow, Ala racemase from *G. stearothermophilus* (PDB code 1NIU) colored in orange and Ala racemase from *E. coli* (PDB code 2RJG) colored in brown. Cl<sup>-</sup> ion and PLP (pyridoxal phosphate) from BsrV are depicted as spheres and sticks respectively. (B) Molecular surface of BsrV with L1 and L2 loops labeled. Active sites are highlighted in white boxes. Right, zoom of the active site with PLP in sticks. (C) Molecular surface of *E. coli* Ala racemase (Alr<sub>Ec</sub>) with L1 and L2 loops labeled. Active sites are highlighted in white boxes. Right, zoom of the active site with PLP in sticks. Panels B and C keep the same orientation. (D)

Electrostatic Potential molecular surface for BsrV and Alr<sub>EC</sub>. Acidic regions are colored in red and basic regions in blue. Active sites are highlighted in white boxes.

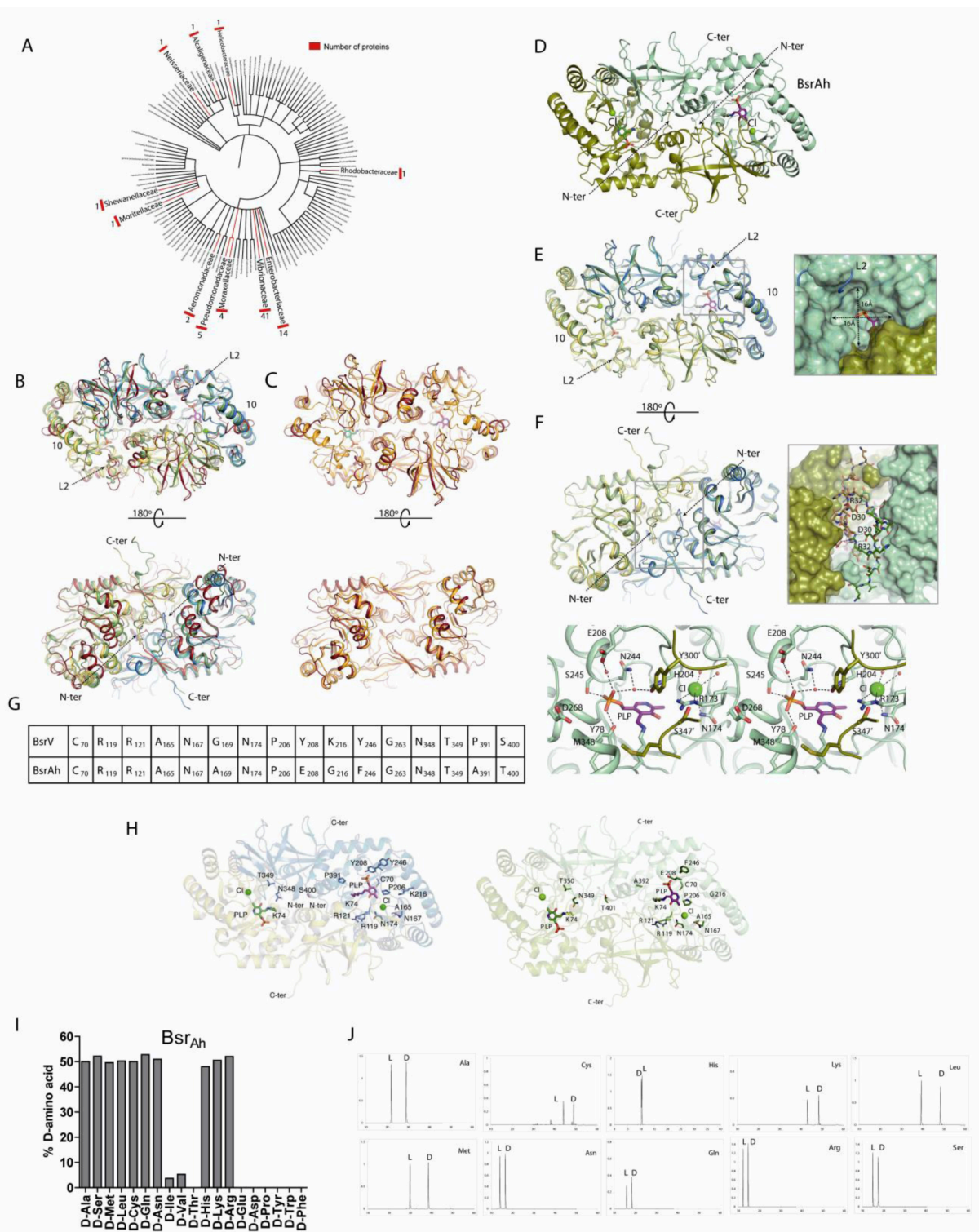


**Supplementary Figure S3.** BsrV and  $\Delta$ Cl-BsrV active site structure and kinetic parameters. Related to Figure 3. (A) Structural environment of Pro25 as part of the N-terminal extension in BsrV. Monomers of BsrV are colored as in Figure 2A. Relevant residues, involved in Pro25 packing, are represented as sticks. (B) Stereo view of the BsrV active site. Monomers of BsrV are colored as in Figure 2A. Relevant residues are

represented in sticks. PLP molecules are drawn in sticks. Cl<sup>-</sup> ions and water molecules as green and red spheres respectively. (C) Left, Stereo view of the structural merge of Alr<sub>Ec</sub> and BsrV active site. In grey is represented the structure of Alr<sub>Ec</sub> and in blue BsrV. Relevant residues of the BsrV catalytic machinery (blue sticks) are labeled. The equivalent residues from Alr<sub>Ec</sub> are represented as grey sticks and its N-carboxylated lysine is labeled in red. The green sphere represents BsrV's chloride atom. PLP molecules are drawn in sticks. Right, stabilization of the chloride ion is shown and distances are labeled. Analysis of the Cl coordination using the Cambridge Crystallographic Data Base (<http://webcsd.ccdc.cam.ac.uk/>) reveals that the average distances for Cl-N and Cl-O are 3.29 (± 0.5) Å and 3.19 (± 0.3) Å respectively. A search for all the deposited protein structures presenting a halogen coordinated by Arg and Asn residues was done by using Relibase

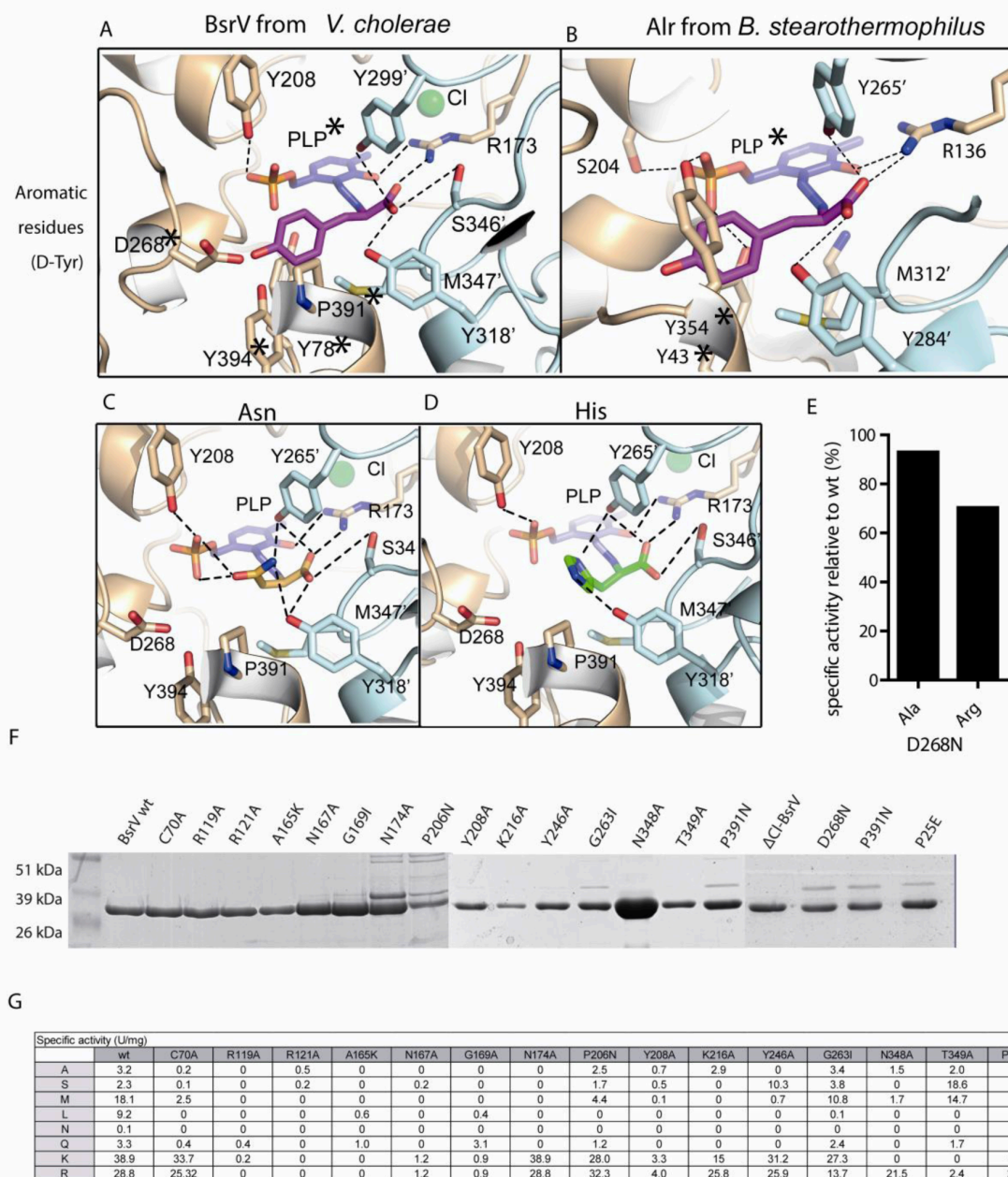
(<http://www.ccdc.cam.ac.uk/Solutions/FreeSoftware/Pages/Relibase.aspx>). In all the 38 entries in PDB presenting such geometry the halogen corresponded to a chloride ion. (D) Representative chromatograms of Marfey's derivatized amino acids (Arg and Lys) after in vitro reaction with apo-BsrV for 90min. (E) Kinetic parameters ( $K_m$ ,  $k_{cat}$  and  $k_{cat}/K_m$ ) for various amounts (1.25-60 mM) of Arg and Lys were calculated using data obtained with 0.47 μM purified apo-BsrV racemase. The enzymatic reactions were measured as described in Materials and Methods. Results in panels B are means ± SD of triplicates from two independent experiments.





**Supplementary Figure S4.** Identification of 16 residues that constitute the molecular footprint for BsrV multispecificity and experimental validation. (A) Phylogenetic tree of the families of 74 proteins identified, belonging to 37 different species. Red leaves correspond to protein families of the proteins in the dataset, the number of proteins belonging to the given family is also specified. (B) Structural superimposition of Alr<sub>Ah</sub> (red cartoon) with BsrV (colored as in Fig. 2A) and Bsr<sub>Ah</sub> (colored in different shades of green). Loop L2 and the  $\alpha$ -helix  $\alpha$ 10, at the entry channel, are labeled. Chloride atoms are represented as spheres. PLP molecules are drawn in sticks. At the bottom, backside (180° rotation) of the dimer is shown. (C) Structural superimposition of Alr<sub>Ah</sub> (red cartoon) with alanine racemase DadX from *Pseudomonas aeruginosa* (PDB code 1RCQ) (orange cartoon). PLP molecules are drawn in sticks. At the bottom, backside (180° rotation) of the dimer is shown. (D) Crystal structure of Bsr<sub>Ah</sub> dimer with monomers depicted in different shades of green. PLP molecules are drawn in sticks and Cl<sup>-</sup> ions as green spheres. (E) Structural superimposition of Bsr<sub>Ah</sub> (colored as in panel D) with BsrV (colored as in panel B). Loop L2 and the  $\alpha$ -helix  $\alpha$ 10, at the entry channel, are labeled. On the right box, differences in their active-site cavities are shown. Bsr<sub>Ah</sub> molecular surface is colored as in panel A and BsrV is represented in blue cartoon. Dimensions at the entrance of Bsr<sub>Ah</sub> cavity are shown by black arrows. (F) Backside (180° rotation) of the dimer. The Bsr<sub>Ah</sub> N-terminal insertion is highlighted in the box. On the right, close-up view with the N-terminal insertion in sticks. At the bottom, stereoview of Bsr<sub>Ah</sub> active site. Relevant residues and PLP are drawn in sticks, water molecules and Cl<sup>-</sup> ions are drawn as red and green spheres, respectively. (G) Comparative analysis of BsrV and Bsr<sub>Ah</sub> footprint residues. (H) Localization of the footprint residues in

BsrV (left) and Bsr<sub>Ah</sub> (right) structural models. (I) Quantitative Bsr<sub>Ah</sub> D-amino acid production spectra. (J) Chromatograms of racemizable amino acids by Bsr<sub>Ah</sub>.



**Supplementary Figure S5.** DAA recognition by BsrV from *V. cholerae* versus Alr from *B. stearotherophilus*. (A) Docking model with the aromatic residue D-Tyr in BsrV (in purple sticks). (B) Docking model with the aromatic amino acid D-Tyr in Alr<sub>Bs</sub> (in purple sticks). (C) Docking model with Asn in BsrV (in yellow sticks). (D) Docking model with



His in BsrV (in yellow sticks). (E) Specific activity of the BsrV D268N mutant relative to wt on Ala and Arg as representative amino acids of a non-charged and a positively charged amino acid. (F) SDS-PAGE analysis of purified proteins after Ni-NTA affinity chromatography purification. (G) Specific activity of 15 point mutants on BsrV's footprint residues.

**Supplementary Table S1. Plasmids used in this study.**

PLASMID	DESCRIPTION	REFERENCE
<b>pET28b (+)</b>	Kan <sup>r</sup> , lacI. Expression of genes in <i>E. coli</i> , dependent on T7 phage RNA polymerase.	Novagen
<b>pET28b</b> <b>AlrV</b>	Kan <sup>r</sup> . AlrV overexpression. 6His N-terminal fusion. (NdeI/EcoRI). DNA fragment ( <i>V. cholerae</i> 's locus VC0372) amplified with FCP598 and FCP597 primers.	(Lam et al., 2009)
<b>pET28b</b> <b>BsrV WT</b>	Kan <sup>r</sup> . BsrV WT overexpression with 6His C-terminal fusion. (NotI/EcoRI). DNA fragment ( <i>V. cholerae</i> 's locus VC1312) amplified with FCP43 and FCP56 primers.	(Lam et al., 2009)
<b>pET28b</b> <b>Alr<sub>Ah</sub></b>	Kan <sup>r</sup> . Alr <sub>Ah</sub> overexpression with 6His C-terminal fusion. (NdeI/EcoRI). DNA fragment ( <i>A. hydrophila</i> 's locus AHA1015) amplified with FCP108 and FCP109 primers.	This study
<b>pET28b</b> <b>ΔCI-BsrV</b>	Kan <sup>r</sup> . BsrV R173N174/AA with C-terminal fusion of 6 histidines (NcoI/NotI). The mutation was introduced with FCP37 and FCP38 primers, replacing RN/AA in BsrV catalytic center.	This study
<b>pET28b</b> <b>BsrV.tev.His</b>	Kan <sup>r</sup> . BsrV WT with Tev protease recognition site before the 6His C-terminal fusion (NcoI/NotI). DNA fragment amplified with FCP56 and FCP140 primers.	This study
<b>pET28</b> <b>BsrV P25E</b>	Kan <sup>r</sup> . BsrV P25E overexpression with 6His C-terminal fusion. (NcoI/NotI). The mutation was introduced with FCP63 and FCP64, replacing P/E in BsrV 25 position.	
<b>pET28b</b> <b>BsrV C70A</b>	Kan <sup>r</sup> . BsrV C70A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP110 and FCP111 primers, replacing C/A in BsrV 70 position.	This study
<b>pET28b</b> <b>BsrV R119A</b>	Kan <sup>r</sup> . BsrV R119A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP112 and FCP113 primers, replacing R/A in BsrV 119 position.	This study
<b>pET28b</b> <b>BsrV R121A</b>	Kan <sup>r</sup> . BsrV R121A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP114 and FCP115 primers, replacing R/A in BsrV 121 position.	This study
<b>pET28b</b> <b>BsrV A165K</b>	Kan <sup>r</sup> . BsrV A165K overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP116 and FCP117 primers, replacing A/K in BsrV 165 position.	This study
<b>pET28b</b> <b>BsrV N167A</b>	Kan <sup>r</sup> . BsrV N167A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP118 and FCP119 primers, replacing N/A in BsrV 167 position.	This study
<b>pET28b</b> <b>BsrV G169I</b>	Kan <sup>r</sup> . BsrV G169I overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP120 and FCP121 primers, replacing G/I in BsrV 169 position.	This study
<b>pET28b</b> <b>BsrV N174A</b>	Kan <sup>r</sup> . BsrV N174A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP122 and FCP123	This study

	primers, replacing N/A in BsrV 174 position.	
<b>pET28b BsrV P206N</b>	Kan <sup>r</sup> . BsrV P206N overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP124 and FCP125 primers, replacing P/N in BsrV 206 position.	This study
<b>pET28b BsrV Y208A</b>	Kan <sup>r</sup> . BsrV Y208A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP126 and FCP127 replacing Y/A in BsrV 208 position.	This study
<b>pET28b BsrV K216A</b>	Kan <sup>r</sup> . BsrV K216A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP128 and FCP129 primers, replacing K/A in BsrV 216 position.	This study
<b>pET28b BsrV Y246A</b>	Kan <sup>r</sup> . BsrV Y246A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP130 and FCP131 primers, replacing Y/A in BsrV 246 position.	This study
<b>pET28b BsrV G263I</b>	Kan <sup>r</sup> . BsrV G263I overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP132 and FCP133 primers, replacing G/I in BsrV 263 position.	This study
<b>pET28b BsrV D268N</b>	Kan <sup>r</sup> . BsrV D268N overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP157 and FCP158 primers, replacing D/N in BsrV 268 position.	This study
<b>pET28b BsrV N348A</b>	Kan <sup>r</sup> . BsrV N348A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP134 and FCP135 primers, replacing N/A in BsrV 348 position.	This study
<b>pET28b BsrV T349A</b>	Kan <sup>r</sup> . BsrV T349A overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP136 and FCP137 primers, replacing T/A in BsrV 349 position.	This study
<b>pET28b BsrV P391N</b>	Kan <sup>r</sup> . BsrV P391N overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP138 and FCP40 primers, replacing P/N in BsrV 391 position.	This study
<b>pET28b BsrV S377N</b>	Kan <sup>r</sup> . BsrV S377N overexpression with 6His C-terminal fusion (NcoI/NotI). The mutation was introduced with FCP139 primers, replacing S/N in BsrV 377 position.	This study
<b>pET28b Bsr<sub>Ah</sub></b>	Kan <sup>r</sup> . BsrAh overexpression. 6His C-terminal fusion. (NcoI/HindIII). DNA fragment ( <i>A. hydrofila</i> 's locus AHA2607) amplified with FCP345 and FCP388 primers.	This study

**Supplementary Table S2. Primers used in this study.**

NAME	SEQUENCE (5'→3')
FCP37	aaaccatggagcagccgcttctcagtcgcaagaag
FCP38	catatcaatgccggccgcgccatgccgccgtcattcagc
FCP40	ctagttattgctcagcggtg
FCP43	aaagcggccgctttcacgtagaaacgtgggttactggtcccc
FCP56	aaaccatggagcagccgcttctcagtcgcaagaag
FCP57	aaatctagatgccaaaggcgatctgtcgacacgatgg
FCP58	gaattactcatcgtaaagcaggtccagcattggcgcagcagagagagaaagaagg
FCP59	ccttcttctctctctgctgcgccaatgctggacctgcatttgacgatgagtaattc
FCP60	ggtttctccaagttggttggtatagcctagtgatgatgatgatggccgctgctaacgcctgctccataataagacatttc
FCP61	gcagcggccatcatcatcatcactaggctatacccaacaactggagaaacc
FCP62	aaatctagatttactcttgagtcacttttagtcggtg
FCP63	gcggatgtgtaattccgcagcagagag
FCP64	ctctctgctgcggaattacacatccgc
FCP106	aaatctagatgggtgaatggtggatgtg
FCP107	aaagaattcccaagaagatcaagcgct
FCP108	aaagaattccagtgcttcttgtgcatg
FCP109	aaatctagagcccagcagatcaccatgc
FCP110	cacaaaattgccgccatcatg
FCP111	catgatggcggcaatttggtg
FCP112	gaactgatcgtgtgctgctctg
FCP113	cagagcgcacagcgatcagttc
FCP114	ccgtgtggcctctgccagcc
FCP115	ggctggcagaggccacacgg
FCP116	gttcacatcaagctgaatgac
FCP117	gtcattcagcttgatgtgaac
FCP118	catcgcgctggctgacggcggc
FCP119	gccgccgtcagccagcgcgatg
FCP120	ctgaatgacatcgcatgggcc
FCP121	ggcccatgccgatgtcattcag
FCP122	catgggccgtgccggcattg
FCP123	caatgccggcacggcccatg
FCP124	cccactcaacaattacaatg
FCP125	cattgtaattgtgaagtggg
FCP126	cttcccgaatgccaatgcgg

---

<b>FCP127</b>	ccgcattggcattcggaag
<b>FCP128</b>	gtgcgcgcagcactggctc
<b>FCP129</b>	gagccagtgtgcgcgcac
<b>FCP130</b>	gcgaactctgccaccgcgctc
<b>FCP131</b>	gagcgcggtggcagagttgc
<b>FCP132</b>	cgtccaggtatcgtgtgttg
<b>FCP133</b>	caaacaacacgatacctggacg
<b>FCP134</b>	catcgatggccaccactgtg
<b>FCP135</b>	cacagtggaggccatcgatg
<b>FCP136</b>	catcgatgaacgccactgtgg
<b>FCP137</b>	ccacagtggcgttcacgatg
<b>FCP138</b>	caacgctgagctgatttcaatgag
<b>FCP139</b>	aaagcgccgcttcacgtagaaacgtgggtattggttc
<b>FCP140</b>	aaagcgccgcttagtgatgatgatgatgagccgctgctggattggaagtacaggtctcttcacgtagaaacgtgggttact ggtcccc
<b>FCP157</b>	gtgtgtttgtaattgccaaccaacc
<b>FCP158</b>	gggttggttgcaaattaccaacaacac

---

**Supplementary Table S3. Analysis of the bacterial species encoding BsrV-orthologues**

Bacterial (specie)	Gram, class, family	Life style	Bsr-SP lengh* (aa)	Reference
<i>Yersinia pseudotuberculosis</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	20	(Kaasch et al., 2012)
<i>Yersinia pestis</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	20/21	(Williamson and Oyston, 2012)
<i>Proteus mirabilis</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	20	(Jacobsen and Shirliff, 2011)
<i>Proteus penneri</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	20	(Rozalski et al., 2007)
<i>Citrobacter sp.</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facult. pathogen	29	(Pepperell et al., 2002)
<i>Providencia rustigianii</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	20	(O'Hara et al., 2000)
<i>Providencia rettgeria</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	21	(O'Hara et al., 2000)
<i>Providencia stuartii</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	24	(O'Hara et al., 2000)
<i>Photobacterium profundum</i>	Gram -, γ -Proteobacteria, Vibrionaceae	Environmental	21	(El-Hajj et al., 2010)
<i>Moritella sp.</i>	Gram -, γ -Proteobacteria, Vibrionaceae	Environmental	25	(DeLong et al., 1997)
<i>Helicobacter hepaticus</i>	Gram -, ε -Proteobacteria, Helicobacteraceae	Facultative pathogen	23	(Murakami, 2012)
<i>Xenorhabdus bovienii</i>	Gram -, γ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	23	(Sugar et al., 2012)

<b><i>Xenorhabdus nematophila</i></b>	Gram -, $\gamma$ -Proteobacteria, Enterobacteriaceae	Facultative pathogen	23	(Herbert and Goodrich-Blair, 2007)
<b><i>Acinetobacter baumanii</i></b>	Gram -, $\gamma$ -Proteobacteria, Moraxellaceae	Facultative pathogen	28	(Howard et al., 2012)
<b><i>Acinetobacter johnsoni</i></b>	Gram -, $\gamma$ -Proteobacteria, Moraxellaceae	Facultative pathogen	24	(Seifert et al., 1993)
<b><i>Aeromonas hydrophyla</i></b>	Gram -, $\gamma$ -Proteobacteria, Aeromonadaceae	Facultative pathogen	22	(Chang et al., 1997)
<b><i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i></b>	Gram -, $\gamma$ -Proteobacteria, Aeromonadaceae	Facultative pathogen	22	(Romer Villumsen et al., 2012)
<b><i>Kingella oralis</i></b>	Gram -, $\beta$ -Proteobacteria, Neisseriaceae	Facultative pathogen	24	(Chen, 1996)
<b><i>Pseudomonas putida</i></b>	Gram -, $\gamma$ -Proteobacteria, Pseudomonadaceae	Facultative pathogen	24	(Kim et al., 2012)
<b><i>Taylorella equigenitalis</i></b>	Gram -, $\beta$ -Proteobacteria, Alcaligenaceae	Facultative pathogen	23	(Timoney, 1996)
<b><i>Roseobacter</i> sp.</b>	Gram -, $\alpha$ -Proteobacteria, Rhodobacteraceae	Environmental	ND	(Lenk et al., 2012)
<b><i>Vibrio cholerae</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Facultative pathogen	23/31	(Harris et al., 2012)
<b><i>Vibrio parahaemolyticus</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Facultative pathogen	25	(Su and Liu, 2007)
<b><i>Vibrio fischeri</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Facultative pathogen	20/21	(Su and Liu, 2007)
<b><i>Vibrio furmisi</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Facultative pathogen	25	(Austin, 2010)
<b><i>Vibrio mimicus</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Facultative pathogen	23	(Austin, 2010)
<b><i>Vibrio caribbenthicus</i></b>	Gram -, $\gamma$ -Proteobacteria, Vibrionaceae	Environmental	23	(Austin, 2010)



<b><i>Vibrio coralliilyticus</i></b>	Gram -, γ-Proteobacteria, Vibrionaceae	Facultative pathogen	23	(Austin, 2010)
<b><i>Vibrio metschnikovii</i></b>	Gram -, γ-Proteobacteria, Vibrionaceae	Facultative pathogen	23	(Austin, 2010)
<b><i>Shewanella halifaxensis</i></b>	Gram -, γ-Proteobacteria, Shewanellaceae	Environmental	22	(Zhao et al., 2006)
<b><i>Reinekea sp.</i></b>	Gram -, γ-Proteobacteria, Reinekea	Environmental	ND	(Pinhassi et al., 2007)

\*Signal peptide length obtained with the SignalP 4.1 Server online tool. ND: Not determined

# Supplementary Table S4. BsrV orthologs identified based on *in silico* comparative analysis.

BsrV orthologs
>gi 170024003 ref YP_001720508.1  alanine racemase [Yersinia pseudotuberculosis YPIII]
>gi 169750537 gb ACA68055.1  alanine racemase [Yersinia pseudotuberculosis YPIII]
>gi 186895776 ref YP_001872888.1  alanine racemase [Yersinia pseudotuberculosis PB1/+]
>gi 186698802 gb ACC89431.1  alanine racemase [Yersinia pseudotuberculosis PB1/+]
>gi 51596715 ref YP_070906.1  alanine racemase [Yersinia pseudotuberculosis IP 32953]
>gi 51589997 emb CAH21630.1  putative alanine racemase [Yersinia pseudotuberculosis IP 32953]
>gi 22125742 ref NP_669165.1  alanine racemase [Yersinia pestis KIM 10]
>gi 108812106 ref YP_647873.1  alanine racemase [Yersinia pestis Nepal516]
>gi 20137378 sp Q8ZFL4.1 ALR2_YERPE RecName: Full=Alanine racemase 2
>gi 145599043 ref YP_001163119.1  alanine racemase [Yersinia pestis Pestoides F]
>gi 149366357 ref ZP_01888391.1  putative alanine racemase [Yersinia pestis CA88-4125]
>gi 153950742 ref YP_001400629.1  alanine racemase [Yersinia pseudotuberculosis IP 31758]
>gi 165925439 ref ZP_02221271.1  alanine racemase [Yersinia pestis biovar Orientalis str. F1991016]
>gi 165938444 ref ZP_02227001.1  alanine racemase [Yersinia pestis biovar Orientalis str. IP275]
>gi 166008560 ref ZP_02229458.1  alanine racemase [Yersinia pestis biovar Antiqua str. E1979001]
>gi 166210808 ref ZP_02236843.1  alanine racemase [Yersinia pestis biovar Antiqua str. B42003004]
>gi 167422631 ref ZP_02314384.1  alanine racemase [Yersinia pestis biovar Orientalis str. MG05-1020]
>gi 167426368 ref ZP_02318121.1  alanine racemase [Yersinia pestis biovar Mediaevalis str. K1973002]
>gi 167468736 ref ZP_02333440.1  alanine racemase [Yersinia pestis FV-1]
>gi 218928815 ref YP_002346690.1  alanine racemase [Yersinia pestis CO92]
>gi 229841680 ref ZP_04461838.1  putative alanine racemase [Yersinia pestis biovar Orientalis str. PEXU2]
>gi 229843797 ref ZP_04463940.1  putative alanine racemase [Yersinia pestis biovar Orientalis str. India 195]
>gi 229894540 ref ZP_04509722.1  putative alanine racemase [Yersinia pestis Pestoides A]
>gi 229902427 ref ZP_04517546.1  putative alanine racemase [Yersinia pestis Nepal516]
>gi 270490394 ref ZP_06207468.1  alanine racemase [Yersinia pestis KIM D27]
>gi 294504124 ref YP_003568186.1  alanine racemase [Yersinia pestis Z176003]
>gi 21958663 gb AAM85416.1 AE013788_2 putative alanine racemase [Yersinia pestis KIM 10]
>gi 108775754 gb ABG18273.1  alanine racemase [Yersinia pestis Nepal516]
>gi 115347426 emb CAL20330.1  putative alanine racemase [Yersinia pestis CO92]
>gi 145210739 gb ABP40146.1  alanine racemase [Yersinia pestis Pestoides F]
>gi 149290731 gb EDM40806.1  putative alanine racemase [Yersinia pestis CA88-4125]
>gi 152962237 gb ABS49698.1  alanine racemase [Yersinia pseudotuberculosis IP 31758]
>gi 165913559 gb EDR32179.1  alanine racemase [Yersinia pestis biovar Orientalis str. IP275]
>gi 165922548 gb EDR39699.1  alanine racemase [Yersinia pestis biovar Orientalis str. F1991016]
>gi 165992942 gb EDR45243.1  alanine racemase [Yersinia pestis biovar Antiqua str. E1979001]
>gi 166207988 gb EDR52468.1  alanine racemase [Yersinia pestis biovar Antiqua str. B42003004]
>gi 166958478 gb EDR55499.1  alanine racemase [Yersinia pestis biovar Orientalis str. MG05-1020]
>gi 167054723 gb EDR64527.1  alanine racemase [Yersinia pestis biovar Mediaevalis str. K1973002]
>gi 229680473 gb EEO076570.1  putative alanine racemase [Yersinia pestis Nepal516]
>gi 229689405 gb EEO81468.1  putative alanine racemase [Yersinia pestis biovar Orientalis str. India 195]
>gi 229694143 gb EEO84191.1  putative alanine racemase [Yersinia pestis biovar Orientalis str. PEXU2]
>gi 229703559 gb EEO90576.1  putative alanine racemase [Yersinia pestis Pestoides A]
>gi 262362240 gb ACY58961.1  alanine racemase [Yersinia pestis D106004]
>gi 262366174 gb ACY62731.1  alanine racemase [Yersinia pestis D182038]
>gi 270338898 gb EFA49675.1  alanine racemase [Yersinia pestis KIM D27]
>gi 294354583 gb ADE64924.1  alanine racemase [Yersinia pestis Z176003]
>gi 320014832 gb ADV98403.1  putative alanine racemase [Yersinia pestis biovar Mediaevalis str. Harbin 35]
>gi 108807830 ref YP_651746.1  alanine racemase [Yersinia pestis Antiqua]
>gi 167401740 ref ZP_02307231.1  alanine racemase [Yersinia pestis biovar Antiqua str. UG05-0454]
>gi 108779743 gb ABG13801.1  putative alanine racemase [Yersinia pestis Antiqua]
>gi 167048845 gb EDR60253.1  alanine racemase [Yersinia pestis biovar Antiqua str. UG05-0454]
>gi 237731612 ref ZP_04562093.1  alanine racemase [Citrobacter sp. 30_2]
>gi 226907151 gb EEH93069.1  alanine racemase [Citrobacter sp. 30_2]
>gi 226327902 ref ZP_03803420.1  hypothetical protein PROPEN_01783 [Proteus penneri ATCC 35198]

>gi|225203606|gb|EEG85960.1| hypothetical protein PROPEN\_01783 [Proteus penneri ATCC 35198]  
 >gi|197285632|ref|YP\_002151504.1| alanine racemase [Proteus mirabilis HI4320]  
 >gi|194683119|emb|CAR43692.1| putative alanine racemase [Proteus mirabilis HI4320]  
 >gi|227356137|ref|ZP\_03840526.1| alanine racemase [Proteus mirabilis ATCC 29906]  
 >gi|227163601|gb|EEI48517.1| alanine racemase [Proteus mirabilis ATCC 29906]  
 >gi|212710795|ref|ZP\_03318923.1| hypothetical protein PROVALCAL\_01863 [Providencia alcalifaciens DSM 30120]  
 >gi|212686492|gb|EEB46020.1| hypothetical protein PROVALCAL\_01863 [Providencia alcalifaciens DSM 30120]  
 >gi|261345933|ref|ZP\_05973577.1| alanine racemase [Providencia rustigianii DSM 4541]  
 >gi|282566018|gb|EFB71553.1| alanine racemase [Providencia rustigianii DSM 4541]  
 >gi|268589404|ref|ZP\_06123625.1| alanine racemase [Providencia rettgeri DSM 1131]  
 >gi|291315242|gb|EFE55695.1| alanine racemase [Providencia rettgeri DSM 1131]  
 >gi|183599355|ref|ZP\_02960848.1| hypothetical protein PROSTU\_02822 [Providencia stuartii ATCC 25827]  
 >gi|188021591|gb|EDU59631.1| hypothetical protein PROSTU\_02822 [Providencia stuartii ATCC 25827]  
 >gi|197334487|ref|YP\_002155497.1| alanine racemase [Vibrio fischeri MJ11]  
 >gi|197315977|gb|ACH65424.1| alanine racemase [Vibrio fischeri MJ11]  
 >gi|59711342|ref|YP\_204118.1| alanine racemase [Vibrio fischeri ES114]  
 >gi|59479443|gb|AAW85230.1| alanine racemase [Vibrio fischeri ES114]  
 >gi|54310452|ref|YP\_131472.1| alanine racemase [Photobacterium profundum SS9]  
 >gi|46914893|emb|CAG21670.1| putative alanine racemase [Photobacterium profundum SS9]  
 >gi|153837983|ref|ZP\_01990650.1| alanine racemase [Vibrio parahaemolyticus AQ3810]  
 >gi|149748591|gb|EDM59450.1| alanine racemase [Vibrio parahaemolyticus AQ3810]  
 >gi|28899622|ref|NP\_799227.1| alanine racemase [Vibrio parahaemolyticus RIMD 2210633]  
 >gi|28807874|dbj|BAC61111.1| putative alanine racemase [Vibrio parahaemolyticus RIMD 2210633]  
 >gi|260364032|ref|ZP\_05776760.1| alanine racemase [Vibrio parahaemolyticus K5030]  
 >gi|260876638|ref|ZP\_05888993.1| alanine racemase [Vibrio parahaemolyticus AN-5034]  
 >gi|260898078|ref|ZP\_05906574.1| alanine racemase [Vibrio parahaemolyticus Peru-466]  
 >gi|260902333|ref|ZP\_05910728.1| alanine racemase [Vibrio parahaemolyticus AQ4037]  
 >gi|308085806|gb|EFO35501.1| alanine racemase [Vibrio parahaemolyticus Peru-466]  
 >gi|308093954|gb|EFO43649.1| alanine racemase [Vibrio parahaemolyticus AN-5034]  
 >gi|308110937|gb|EFO48477.1| alanine racemase [Vibrio parahaemolyticus AQ4037]  
 >gi|308114571|gb|EFO52111.1| alanine racemase [Vibrio parahaemolyticus K5030]  
 >gi|167624749|ref|YP\_001675043.1| alanine racemase [Shewanella halifaxensis HAW-EB4]  
 >gi|167354771|gb|ABZ77384.1| alanine racemase [Shewanella halifaxensis HAW-EB4]  
 >gi|312882516|ref|ZP\_07742257.1| alanine racemase [Vibrio caribbenthicus ATCC BAA-2122]  
 >gi|309369916|gb|EFP97427.1| alanine racemase [Vibrio caribbenthicus ATCC BAA-2122]  
 >gi|260778604|ref|ZP\_05887496.1| alanine racemase [Vibrio coralliilyticus ATCC BAA-450]  
 >gi|260604768|gb|EEX31063.1| alanine racemase [Vibrio coralliilyticus ATCC BAA-450]  
 >gi|260772889|ref|ZP\_05881805.1| alanine racemase [Vibrio metschnikovii CIP 69.14]  
 >gi|260612028|gb|EEX37231.1| alanine racemase [Vibrio metschnikovii CIP 69.14]  
 >gi|149910818|ref|ZP\_01899452.1| alanine racemase [Moritella sp. PE36]  
 >gi|149806152|gb|EDM66132.1| alanine racemase [Moritella sp. PE36]  
 >gi|54303604|ref|YP\_133597.1| alanine racemase [Photobacterium profundum SS9]  
 >gi|46917035|emb|CAG23797.1| putative alanine racemase [Photobacterium profundum SS9]  
 >gi|27366912|ref|NP\_762439.1| alanine racemase [Vibrio vulnificus CMCP6]  
 >gi|34098524|sp|Q8D6Q0.1|ALR2\_VIBVU RecName: Full=Alanine racemase 2  
 >gi|27358479|gb|AAO07429.1| alanine racemase [Vibrio vulnificus CMCP6]  
 >gi|37676689|ref|NP\_937085.1| alanine racemase [Vibrio vulnificus YJ016]  
 >gi|37201232|dbj|BAC97055.1| alanine racemase [Vibrio vulnificus YJ016]  
 >gi|320158793|ref|YP\_004191171.1| alanine racemase [Vibrio vulnificus MO6-24/O]  
 >gi|319934105|gb|ADV88968.1| alanine racemase [Vibrio vulnificus MO6-24/O]  
 >gi|161582006|ref|NP\_230956.2| alanine racemase [Vibrio cholerae O1 biovar El Tor str. N16961]  
 >gi|255745708|ref|ZP\_05419656.1| alanine racemase [Vibrio cholera CIRS 101]  
 >gi|262159017|ref|ZP\_06030129.1| alanine racemase [Vibrio cholerae INDRE 91/1]  
 >gi|262169374|ref|ZP\_06037066.1| alanine racemase [Vibrio cholerae RC27]  
 >gi|255736783|gb|EET92180.1| alanine racemase [Vibrio cholera CIRS 101]  
 >gi|262022187|gb|EEY40896.1| alanine racemase [Vibrio cholerae RC27]  
 >gi|262029202|gb|EEY47854.1| alanine racemase [Vibrio cholerae INDRE 91/1]  
 >gi|121586566|ref|ZP\_01676352.1| alanine racemase, putative [Vibrio cholerae 2740-80]  
 >gi|121726850|ref|ZP\_01680062.1| alanine racemase, putative [Vibrio cholerae V52]

>gi|147673980|ref|YP\_001216876.1| alanine racemase [Vibrio cholerae O395]  
 >gi|153819283|ref|ZP\_01971950.1| alanine racemase, putative [Vibrio cholerae NCTC 8457]  
 >gi|153822911|ref|ZP\_01975578.1| alanine racemase, putative [Vibrio cholerae B33]  
 >gi|227081483|ref|YP\_002810034.1| alanine racemase, putative [Vibrio cholerae M66-2]  
 >gi|229505102|ref|ZP\_04394612.1| alanine racemase [Vibrio cholerae BX 330286]  
 >gi|229511228|ref|ZP\_04400707.1| alanine racemase [Vibrio cholerae B33]  
 >gi|229518346|ref|ZP\_04407790.1| alanine racemase [Vibrio cholerae RC9]  
 >gi|229608106|ref|YP\_002878754.1| alanine racemase [Vibrio cholerae MJ-1236]  
 >gi|254848434|ref|ZP\_05237784.1| alanine racemase [Vibrio cholerae MO10]  
 >gi|298498598|ref|ZP\_07008405.1| alanine racemase [Vibrio cholerae MAK 757]  
 >gi|9655801|gb|AAF94470.1| alanine racemase, putative [Vibrio cholerae O1 biovar El Tor str. N16961]  
 >gi|121549245|gb|EAX59277.1| alanine racemase, putative [Vibrio cholerae 2740-80]  
 >gi|121630754|gb|EAX63139.1| alanine racemase, putative [Vibrio cholerae V52]  
 >gi|126510186|gb|EAX72780.1| alanine racemase, putative [Vibrio cholerae NCTC 8457]  
 >gi|126519563|gb|EAX76786.1| alanine racemase, putative [Vibrio cholerae B33]  
 >gi|146315863|gb|ABQ20402.1| putative alanine racemase [Vibrio cholerae O395]  
 >gi|227009371|gb|ACP05583.1| alanine racemase, putative [Vibrio cholerae M66-2]  
 >gi|227013229|gb|ACP09439.1| alanine racemase, putative [Vibrio cholerae O395]  
 >gi|229345061|gb|EEO10035.1| alanine racemase [Vibrio cholerae RC9]  
 >gi|229351193|gb|EEO16134.1| alanine racemase [Vibrio cholerae B33]  
 >gi|229357325|gb|EEO22242.1| alanine racemase [Vibrio cholerae BX 330286]  
 >gi|229370761|gb|ACQ61184.1| alanine racemase [Vibrio cholerae MJ-1236]  
 >gi|254844139|gb|EET22553.1| alanine racemase [Vibrio cholerae MO10]  
 >gi|297542931|gb|EFH78981.1| alanine racemase [Vibrio cholerae MAK 757]  
 >gi|20137517|sp|Q9KSE5.2|ALR2\_VIBCH RecName: Full=Alanine racemase 2  
 >gi|254226880|ref|ZP\_04920449.1| alanine racemase [Vibrio cholerae V51]  
 >gi|125620597|gb|EAX48962.1| alanine racemase [Vibrio cholerae V51]  
 >gi|297578895|ref|ZP\_06940823.1| conserved hypothetical protein [Vibrio cholerae RC385]  
 >gi|297536489|gb|EFH75322.1| conserved hypothetical protein [Vibrio cholerae RC385]  
 >gi|262190944|ref|ZP\_06049158.1| alanine racemase [Vibrio cholerae CT 5369-93]  
 >gi|262033181|gb|EEY51705.1| alanine racemase [Vibrio cholerae CT 5369-93]  
 >gi|153216974|ref|ZP\_01950738.1| alanine racemase, putative [Vibrio cholerae 1587]  
 >gi|124114006|gb|EAY32826.1| alanine racemase, putative [Vibrio cholerae 1587]  
 >gi|229515688|ref|ZP\_04405147.1| alanine racemase [Vibrio cholerae TMA 21]  
 >gi|229347457|gb|EEO12417.1| alanine racemase [Vibrio cholerae TMA 21]  
 >gi|254286591|ref|ZP\_04961547.1| alanine racemase, putative [Vibrio cholerae AM-19226]  
 >gi|150423349|gb|EDN15294.1| alanine racemase, putative [Vibrio cholerae AM-19226]  
 >gi|153802661|ref|ZP\_01957247.1| alanine racemase, putative [Vibrio cholerae MZO-3]  
 >gi|124121789|gb|EAY40532.1| alanine racemase, putative [Vibrio cholerae MZO-3]  
 >gi|229521424|ref|ZP\_04410843.1| alanine racemase [Vibrio cholerae TM 11079-80]  
 >gi|229341522|gb|EEO06525.1| alanine racemase [Vibrio cholerae TM 11079-80]  
 >gi|153830040|ref|ZP\_01982707.1| putative alanine racemase [Vibrio cholerae 623-39]  
 >gi|148874480|gb|EDL72615.1| putative alanine racemase [Vibrio cholerae 623-39]  
 >gi|153827222|ref|ZP\_01979889.1| alanine racemase, putative [Vibrio cholerae MZO-2]  
 >gi|149738872|gb|EDM53204.1| alanine racemase, putative [Vibrio cholerae MZO-2]  
 >gi|229525903|ref|ZP\_04415308.1| alanine racemase [Vibrio cholerae bv. albensis VL426]  
 >gi|229339484|gb|EEO04501.1| alanine racemase [Vibrio cholerae bv. albensis VL426]  
 >gi|229529606|ref|ZP\_04418996.1| alanine racemase [Vibrio cholerae 12129(1)]  
 >gi|229333380|gb|EEN98866.1| alanine racemase [Vibrio cholerae 12129(1)]  
 >gi|261210961|ref|ZP\_05925251.1| alanine racemase [Vibrio sp. RC341]  
 >gi|260839936|gb|EEX66536.1| alanine racemase [Vibrio sp. RC341]  
 >gi|262402336|ref|ZP\_06078897.1| alanine racemase [Vibrio sp. RC586]  
 >gi|262351118|gb|EEZ00251.1| alanine racemase [Vibrio sp. RC586]  
 >gi|262165566|ref|ZP\_06033303.1| alanine racemase [Vibrio mimicus VM223]  
 >gi|262025282|gb|EEY43950.1| alanine racemase [Vibrio mimicus VM223]  
 >gi|262171654|ref|ZP\_06039332.1| alanine racemase [Vibrio mimicus MB-451]  
 >gi|261892730|gb|EEY38716.1| alanine racemase [Vibrio mimicus MB-451]  
 >gi|258621589|ref|ZP\_05716622.1| Alanine racemase [Vibrio mimicus VM573]  
 >gi|258586207|gb|EEW10923.1| Alanine racemase [Vibrio mimicus VM573]  
 >gi|258627600|ref|ZP\_05722377.1| Alanine racemase [Vibrio mimicus VM603]  
 >gi|258580182|gb|EEW05154.1| Alanine racemase [Vibrio mimicus VM603]

>gi|315181783|gb|ADT88696.1| alanine racemase [Vibrio furnissii NCTC 11218]  
 >gi|260769245|ref|ZP\_05878178.1| alanine racemase [Vibrio furnissii CIP 102972]  
 >gi|260614583|gb|EEX39769.1| alanine racemase [Vibrio furnissii CIP 102972]  
 >gi|28900856|ref|NP\_800511.1| alanine racemase [Vibrio parahaemolyticus RIMD 2210633]  
 >gi|260362532|ref|ZP\_05775452.1| alanine racemase [Vibrio parahaemolyticus K5030]  
 >gi|260879223|ref|ZP\_05891578.1| alanine racemase [Vibrio parahaemolyticus AN-5034]  
 >gi|260896758|ref|ZP\_05905254.1| alanine racemase [Vibrio parahaemolyticus Peru-466]  
 >gi|34098472|sp|Q87HG4.1|ALR2\_VIBPA RecName: Full=Alanine racemase 2  
 >gi|28809302|dbj|BAC62344.1| putative alanine racemase [Vibrio parahaemolyticus RIMD 2210633]  
 >gi|308087318|gb|EFO37013.1| alanine racemase [Vibrio parahaemolyticus Peru-466]  
 >gi|308094088|gb|EFO43783.1| alanine racemase [Vibrio parahaemolyticus AN-5034]  
 >gi|308115278|gb|EFO52818.1| alanine racemase [Vibrio parahaemolyticus K5030]  
 >gi|153836908|ref|ZP\_01989575.1| alanine racemase [Vibrio parahaemolyticus AQ3810]  
 >gi|149749866|gb|EDM60611.1| alanine racemase [Vibrio parahaemolyticus AQ3810]  
 >gi|260901465|ref|ZP\_05909860.1| alanine racemase [Vibrio parahaemolyticus AQ4037]  
 >gi|308107691|gb|EFO45231.1| alanine racemase [Vibrio parahaemolyticus AQ4037]  
 >gi|262395354|ref|YP\_003287207.1| alanine racemase [Vibrio sp. Ex25]  
 >gi|262338948|gb|ACY52742.1| alanine racemase [Vibrio sp. Ex25]  
 >gi|91224603|ref|ZP\_01259864.1| alanine racemase [Vibrio alginolyticus 12G01]  
 >gi|91190491|gb|EAS76759.1| alanine racemase [Vibrio alginolyticus 12G01]  
 >gi|32265965|ref|NP\_859997.1| alanine racemase [Helicobacter hepaticus ATCC 51449]  
 >gi|32262014|gb|AAP77063.1| alanine racemase [Helicobacter hepaticus ATCC 51449]  
 >gi|290476786|ref|YP\_003469697.1| Alanine racemase 2 [Xenorhabdus bovienii SS-2004]  
 >gi|289176130|emb|CBJ82935.1| Alanine racemase 2 [Xenorhabdus bovienii SS-2004]  
 >gi|300724984|ref|YP\_003714309.1| putative alanine racemase [Xenorhabdus nematophila ATCC 19061]  
 >gi|297631526|emb|CBJ92233.1| putative alanine racemase [Xenorhabdus nematophila ATCC 19061]  
 >gi|262368964|ref|ZP\_06062293.1| alanine racemase [Acinetobacter johnsonii SH046]  
 >gi|262316642|gb|EEY97680.1| alanine racemase [Acinetobacter johnsonii SH046]  
 >gi|293608485|ref|ZP\_06690788.1| conserved hypothetical protein [Acinetobacter sp. SH024]  
 >gi|292829058|gb|EFF87420.1| conserved hypothetical protein [Acinetobacter sp. SH024]  
 >gi|193077032|gb|ABO11785.2| alanine racemase [Acinetobacter baumannii ATCC 17978]  
 >gi|126641403|ref|YP\_001084387.1| alanine racemase [Acinetobacter baumannii ATCC 17978]  
 >gi|145298692|ref|YP\_001141533.1| alanine racemase [Aeromonas salmonicida subsp. salmonicida A449]  
 >gi|142851464|gb|ABO89785.1| alanine racemase [Aeromonas salmonicida subsp. salmonicida A449]  
 >gi|117619060|ref|YP\_857116.1| alanine racemase [Aeromonas hydrophila subsp. hydrophila ATCC 7966]  
 >gi|117560467|gb|ABK37415.1| alanine racemase [Aeromonas hydrophila subsp. hydrophila ATCC 7966]  
 >gi|238022778|ref|ZP\_04603204.1| hypothetical protein GCWU000324\_02689 [Kingella oralis ATCC 51147]  
 >gi|237865981|gb|EEP67117.1| hypothetical protein GCWU000324\_02689 [Kingella oralis ATCC 51147]  
 >gi|223939404|ref|ZP\_03631282.1| alanine racemase [bacterium Ellin514]  
 >gi|223891896|gb|EEF58379.1| alanine racemase [bacterium Ellin514]  
 >gi|319779710|ref|YP\_004130623.1| Alanine racemase [Taylorella equigenitalis MCE9]  
 >gi|317109734|gb|ADU92480.1| Alanine racemase [Taylorella equigenitalis MCE9]  
 >gi|73763191|gb|AAZ83975.1| alanine racemase [Pseudomonas putida]  
 >gi|167033190|ref|YP\_001668421.1| alanine racemase [Pseudomonas putida GB-1]  
 >gi|166859678|gb|ABY98085.1| alanine racemase [Pseudomonas putida GB-1]  
 >gi|313498322|gb|ADR59688.1| Alr [Pseudomonas putida BIRD-1]  
 >gi|26990430|ref|NP\_745855.1| alanine racemase [Pseudomonas putida KT2440]  
 >gi|24985399|gb|AAN69319.1|AE016569\_3 alanine racemase, putative [Pseudomonas putida KT2440]  
 >gi|148547263|ref|YP\_001267365.1| alanine racemase [Pseudomonas putida F1]  
 >gi|148511321|gb|ABQ78181.1| alanine racemase [Pseudomonas putida F1]  
 >gi|86138387|ref|ZP\_01056961.1| alanine racemase [Roseobacter sp. MED193]  
 >gi|85824912|gb|EAQ45113.1| alanine racemase [Roseobacter sp. MED193]  
 >gi|88800370|ref|ZP\_01115936.1| alanine racemase [Reinekea sp. MED297]  
 >gi|88776947|gb|EAR08156.1| alanine racemase [Reinekea sp. MED297]

## **Supplementary Materials and Methods**

**Growth conditions.** All *V. cholerae* strains used in this study were derived from the sequenced El Tor clinical isolate N16961 (Heidelberg et al., 2000). *E. coli* and *V. cholerae* were grown in Luria broth (LB), on LB agar plates or in minimal medium ( $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$  64 g/L,  $\text{KH}_2\text{PO}_4$  15 g/L,  $\text{NH}_4\text{Cl}$  5 g/L,  $\text{NaCl}$  2.5 g/L per liter of distilled water) supplemented with glucose 0.2% at 37°C. Antibiotics were used at the following concentrations (per ml): 200 µg streptomycin (Sm); 50 µg ampicillin (Ap) or carbenicillin (Cb) and 50 µg kanamycin (Kn) (*V. cholerae*).

**DNA amplification and cloning.** Plasmids and primers are listed in Tables S1 and S2. DNA was amplified using 1 U·ml<sup>-1</sup> of DNA polymerase from *Pyrococcus furiosus* (Roche Molecular Biochemicals) in its recommended buffer with  $\text{MgCl}_2$  3 mM, dNTP mixture 0.5 mM and 50 pmol of each primer (Sigma-Aldrich). When required, overlap extension techniques were used following appropriate protocols (Nelson and Fitch, 2011).

**Bacterial transformation.** *E. coli* competence was induced following Inoue's method (Inoue et al., 1990). The transformation was carried out following the method described by Hanahan (Hanahan, 1985).

**Protein expression and purification.** The *V. cholerae* genes encoding wt and mutant BsrV, and AlrV, Alr<sub>Ah</sub> and Bsr<sub>Ah</sub> were cloned in pET28b (Novagen) for expression in *E. coli* BL21(DE3) cells (Rosenberg et al., 1987). Expression was induced (at OD<sub>600</sub>=0.4) with 1 mM IPTG for 3 h. Cell pellets were resuspended in 50 mM Tris HCl pH 7.2, 150 mM NaCl, 10% glycerol, and Complete Protease Inhibitor Cocktail Tablets (Roche), and lysed with 3 passes through a French press. Proteins were purified from cleared lysates (30 min, 50,000 rpm) on Ni-NTA agarose columns (Qiagen), and eluted with a

discontinuous imidazol gradient. Pure proteins were visualized by SDS-PAGE electrophoretic protein separation (Laemmli and Favre, 1973). N-terminal sequence revealed that BsrV and Bsr<sub>Ah</sub> were purified as the predicted full length mature proteins: BsrV (residues 24-407); Bsr<sub>Ah</sub> (residues 22-408). Untagged BsrV was purified from its His-tagged derivative (see Table S2) which presents a Tobacco etch virus (TEV) protease cleavage site preceding the His-tag. TEV protease (Sigma) digestion was performed at 30°C for 6h, in 25 mM Tris-HCl, pH 8.0, 150 - 500 mM NaCl, 14 mM  $\beta$ -mercaptoethanol.

**Racemase activity assays.** All racemase assays were done in 2 fundamental steps. First, the mixture was prepared in a 50  $\mu$ l final volume with: Tris-HCl 50 mM pH 8, L-amino acid at various concentrations, and 0.47  $\mu$ M of racemase (AlrV, BsrV,  $\Delta$ CI-BsrV, Bsr<sub>Ah</sub> and all single BsrV-mutants in the SDP signature residues). All amino acids used as substrate were purchased from Sigma Aldrich (St. Louis, MO). The reaction was performed for 5 minutes (steady-state kinetics) or 90 minutes (long term assay). Then reactions were stopped by boiling during 10 min and samples were spin for 10 minutes at 14,500 rpm to remove inactivated proteins. Next, D-amino acid production was determined by coupling 10  $\mu$ l of the extract into 150  $\mu$ l of a reaction containing: sodium phosphate buffer 100 mM pH 7, *Trigonopsis variabilis* D-amino acid oxidase (DAAO) (Komarova et al., 2012) 3.6 U/ml, horseradish peroxidase 1 U/ml, o-phenilendiamine (OPD) 2 mg/ml and FAD 3 mg/ml. This two step assay permits the quantification of H<sub>2</sub>O<sub>2</sub> (DAAO is able to produce  $\alpha$ -ketoacid, NH<sub>3</sub> and H<sub>2</sub>O<sub>2</sub> from DAA). Peroxidase reduces H<sub>2</sub>O<sub>2</sub> releasing free O<sub>2</sub> that reacts with OPD, leading to the production of 2,3-

diaminophenazine. The reaction was incubated for 2 hours at 37°C and inactivated with HCl 3 M, giving a colorimetric product that can be measured at 492 nm.

For those cases where DAAO activity was a constrain (i.e. His, Cys), or for competition assays, Marfey's (FDAA)-derivatization of L- and D-forms of amino acids and HPLC analysis was followed (described below).

The kinetic parameters were determined by measuring production rates for different substrate concentrations (1.25 – 60 mM) of each L-amino acid with 0.47  $\mu$ M racemase. The activities for each substrate concentration were measured thrice. All mean activities were plotted and adjusted to a Michaelis–Menten model to determine the apparent  $K_m$ ,  $V_{max}$ , and  $k_{cat}$  by non-linear regression (Microsoft Office 2007 and Solver macro) (De Levie, 2001).  $k_{cat}$  was determined as  $V_{max}/[E_0]$ , where  $[E_0]$  = nmol of racemase/ml. Reaction kinetics for His and Cys were not established given that L-His is not a DAAO substrate and L-Cys is a peroxidase inhibitor. The  $\epsilon$  used to determine the kinetic parameters was 1,220 m/cm and the molecular weight of the enzyme monomer without signal peptide used to determine the  $k_{cat}$  was 42 kDa.

**Marfey-derivatized amino acids analysis.** The product from a racemization reaction was derivatized with L-FDAA (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide, Marfey's reagent, Thermo Scientific). First, an equal volume of NaHCO<sub>3</sub> 0.5 M was added to the racemization reaction; then, 6  $\mu$ l of this reaction was reacted with FDAA (10 mg/ml in acetone) at 80°C for 3 min. The reaction was stopped with HCl 2N and the samples were filtered. The products were separated with a linear gradient of triethylamine phosphate/acetonitrile in HPLC with an Aeris peptide column (250 x 4.6 mm; 3.6  $\mu$ m particle size) (Phenomenex, USA) and detected at Abs.340 nm.



**Crystallization of BsrV,  $\Delta$ CI-BsrV, Bsr<sub>Ah</sub> and Alr<sub>Ah</sub>.** High-throughput techniques using a NanoDrop™ ExtY robot (Innovadyne Technologies Inc.) were carried out to test initial crystallization assays based on commercial Qiagen® screenings, JCSG+Suite™ and PACT™ Suite, and Hampton Research® screenings, IINDEX™ and Crystal Screen™ 1 and 2. All crystallization nano-trials were performed by sitting-drop vapour-diffusion method at 291 K using Innovaplate™ SD-2 microplates (Innovadyne Technologies Inc.) and mixing equal amounts (250 nl) of purified aliquots of BsrV (4.5 mg/ml and 3.0 mg/ml, both in 20 mM Tris, pH 7.0), BsrV-RN/AA ( $\Delta$ CI-BsrV) (6.25 mg/ml and 3.5 mg/ml, both in 20 mM Tris pH 7.0), Bsr<sub>Ah</sub> (8 mg/ml in 20 mM Tris pH 7.0), Alr<sub>Ah</sub> (6 mg/ml in 20 mM Tris–HCl pH 7.5) and precipitant solutions. NanoDrop robot drops (500 nl, protein and precipitant solutions) were equilibrated against 65 ml of mother liquor. BsrV rhomboid microcrystals grew from several conditions of INDEX™, PACT™ Suite and JCSG+Suite™. The conditions were systematically improved, and finally, the best quality crystals grew in the condition composed by 0.1 M Bis-Tris propane pH 7.5, 0.2 M Sodium Iodide, and 20% (p/v) of PEG 3350. These crystals were obtained using the sitting drop method (1:1 volume proportion), mixing 2  $\mu$ l of wt BsrV (holoenzyme) at 3.0 mg/ml with 1  $\mu$ l of precipitant solution, and equilibrated against 500  $\mu$ l of mother liquor. Good quality crystals from purified  $\Delta$ CI-BsrV (3.5 mg/ml in Tris pH 7.0) were obtained in the improved condition constituted by 0.1 M Bis-Tris Propane pH 7.5, 0.2 M Sodium Bromide and 29% PEG 3,350. Bsr<sub>Ah</sub> microcrystals grew from JCSG+Suite™ condition number 77 (5mM CoCl<sub>2</sub>, 5mM CdCl<sub>2</sub>, 5mM MgCl<sub>2</sub>, 5mM NiCl<sub>2</sub>, 0.1M Hepes pH 7.5 and 12% PEG 3350). The Bsr<sub>Ah</sub> crystallization condition was scaled to a 3  $\mu$ l drop (precipitant/protein ratio 2:1) and finally improved to obtain good quality diffracting

crystals at the condition composed by 5 mM CoCl<sub>2</sub>, 5 mM CdCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 5 mM NiCl<sub>2</sub>, 0.1 M Hepes pH 7.5 and 16% PEG 3350. Alr<sub>Ah</sub> microcrystals grew from the PACT™ Suite condition number 63 (0.1 M Bis Tris propane pH 6.5, 0.2 M Sodium iodide, 20% PEG 3350). Various additives and detergents (Hampton Research) were tested around these conditions, but none of them led to improvement of the crystals size and/or diffraction quality. The final condition (0.1 M Bis Tris propane pH 6.5, 0.2 M Sodium iodide, 14% PEG 3350) was scaled to bigger drops (2 µl).

**X-ray data collection and structural determination.** Holoenzyme and  $\Delta$ CI-BsrV crystals of BsrV were soaked for 5 s in a cryoprotectant solution (15% v/v of glycerol diluted in crystallization mother liquor) prior to flash-cooling at 100 K in liquid nitrogen. X-ray data sets up to 1.1 Å resolution were collected using synchrotron radiation facility at the beamline ID29 (ESRF, Grenoble) for the holoenzyme and several data sets up to 1.5 Å for the  $\Delta$ CI-BsrV crystals at the X06SA beamline (SLS, Villigen). Both holoenzyme and  $\Delta$ CI-BsrV data sets were collected at 100 K using a PILATUS 6M detector (holoenzyme: oscillation range 0.1°,  $\Delta$ CI-BsrV: oscillation range 0.25°). All data sets were processed using XDS (Kabsch, 1993) and scaled using SCALA (Evans, 2006) from CCP4 program suite. Data processing results are summarized in Table 1. BsrV structure was solved by molecular replacement method (MR) with the program MolRep from CCP4 program suite using the *E. coli* alanine racemase (PDB code 2RJG) as template. The structural model was refined with the PHENIX program (Adams et al., 2010) and manually improved using COOT (Emsley and Cowtan, 2004). The final models for both holoenzyme and  $\Delta$ CI-BsrV presented good stereochemistry. Crystals of Alr<sub>Ah</sub> were soaked for 5 s in a cryoprotectant solution (15% v/v of glycerol diluted in

crystallization mother liquor) prior to flash-cooling at 100 K in liquid nitrogen. X-ray data sets up to 1.4 Å resolution were collected using synchrotron radiation facility at the beamline ID14-4 (ESRF, Grenoble) using an ADSC Q315r detector (oscillation range 1°), and processed (results are summarized in Table 1) using XDS (Kabsch, 1993) and scaled using SCALA (Evans, 2006) from CCP4 program suite. Bsr<sub>Ah</sub> structure was solved by MR method with the program MolRep from CCP4 program suite using BsrV structure as template. The structural model of Bsr<sub>Ah</sub> was refined with the PHENIX program (Adams et al., 2010) and manually improved using COOT (Emsley and Cowtan, 2004). Crystals of Alr<sub>Bs</sub> were cryoprotected by a quick soak in 0.1 M Bis Tris propane pH 6.5, 0.2 M Sodium iodide, 14% PEG 3350, 15% glycerol. X-ray diffraction were collected at the ID29 beamline (ESRF, Grenoble) using a PILATUS 6M detector (oscillation range 0.1°). Although several X-ray data collections were measured, the maximum resolution reached was 3.25 Å. Diffraction data were processed (results are summarized in Table 1) using XDS (Kabsch, 1993) and scaled using SCALA (Evans, 2006) from CCP4 program suite. The crystal structure of Alr<sub>Ah</sub> was solved by MR method, using the BALBES program (Long et al., 2008). A very incomplete initial model was obtained (<50% of the residues in the AU were correctly positioned, R-factor 47%) and subsequently, the main chain was fixed in density using strict geometry restraints and the incomplete regions were traced manually in COOT (Emsley and Cowtan, 2004). adding alanines as first attempt. Due to the low resolution, during this process the alanine racemase structure from *Pseudomonas aeruginosa* (PDB code 1RCQ) was used as template. The structure was refined using BUSTER-TNT program (Bricogne, 2011) with NCS, TLS, and automatic low resolution restraints. The B-factors were

refined as groups. Several cycles of refinement and manual model rebuilding were necessary to reach a good model, which included four independent proteins of 357 amino acids. The waters molecules were assigned with BUSTER-TNT. The refinement statistics are summarized in Table 1. The stereochemistry of the models was verified with MOLPROBITY (Chen et al., 2009).

### **Identification of new putative broad-spectrum racemases and the residues responsible for the substrate specificity**

The initial retrieval of PLP-racemases sequences for studying the subfamily structure was performed by searching with BLAST (Altschul et al., 1997), using AlrV as query sequence, against a non-redundant protein database, with a restrictive E-value cutoff of  $1e^{-10}$ . This search resulted in 2,967 sequences, which included BsrV. Protein sequences not present in the dataset due to the redundancy filter for which experimental or structural information regarding substrate specificity was available were added, including alanine racemases from *E. coli*, *V. cholerae* and *A. hydrophila*. The resulting sequences were filtered selecting only those mapped into the UniProt database, yielding a preliminary set of 1,355 protein sequences.

The protein dataset was aligned using the multiple sequence alignment tool MUSCLE6 (v3.8.31). The resulting MSA was filtered out for empty columns and for redundant proteins (using a 95% redundancy threshold) using the Jalview7 tool (v2.7), resulting in 137 sequences in the final MSA used for inferring the PLP-racemase subfamily composition and associated “specificity determining positions” (SDPs).

Identification of SDPs was carried out using the Xdet and S3Det programs implemented in the JDet package. JDet allows extracting, visualizing and manipulating fully

conserved positions and family dependent positions in MSAs<sup>3</sup>. S3Det not only detects SDPs but also defines the subfamily composition of the MSA<sup>2</sup>.

A phylogenetic tree was constructed from the same MSA to compare it with the partition of the MSA into subfamilies generated by S3Det. The tree was based on the neighbor joining method using substitution matrix BLOSUM62 and was built using the Jalview7 tool (v2.7). The manipulation and display of the resulting phylogenetic tree was carried out with iTool<sup>8</sup>.

After this initial filtered alignment was used for predicting the structure of the PLP-racemase family and the associated SDPs, a comprehensive search and classification of this family of proteins was carried out. Using AlrV and BsrV as query sequences, two BLAST searches were carried out (E-value cutoff  $1e^{-5}$ ) yielding a final joint set of 2,540 sequences. The two searches converged virtually to the same set of sequences. These sequences were aligned with MUSCLE (Edgar, 2004) and S3Det was used to infer the subfamily composition of the resulting MSA. The resulting subfamilies were assigned to AlrV-like or BsrV-like (“specific” and “broad”) depending on their “enrichment” in the sequences previously classified (in the MSA of 137 sequences). Consequently, a subfamily of 1,800 sequences containing 68 out of the 84 AlrV-like (and only 1 out of the 11 BsrV-like) was assigned to the “specific” class. The proteins within a cluster of 74 sequences containing 10 out of the 11 BsrV-like (and no AlrV-like) were consequently assigned as putative new BsrV-like.

**Supplementary references**

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). *Nucleic Acids Res* **25**, 3389-3402.
- Austin, B. (2010). *Vet Microbiol* **140**, 310-317.
- Bricogne, G.B.E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., *et al.* (2011). BUSTER version X.Y.Z. Cambridge, United Kingdom: Global Phasing Ltd.
- Chang, C.Y., Thompson, H., Rodman, N., Bylander, J., and Thomas, J. (1997). *Ann Clin Lab Sci* **27**, 254-259.
- Chen, C. (1996). *Oral Microbiol Immunol* **11**, 425-427.
- Chen, Y., Zhang, W., Shi, Q., Hesek, D., Lee, M., Mobashery, S., and Shoichet, B.K. (2009). *J Am Chem Soc* **131**, 14345-14354.
- De Levie, R. (2001). *How to Use Excel in Analytical Chemistry and in General Scientific Data Analysis*. Cambridge Univ Press, Cambridge.
- Delong, E.F., Franks, D.G., and Yayanos, A.A. (1997). *Appl Environ Microbiol* **63**, 2105-2108.
- Edgar, R.C. (2004). *Nucleic Acids Res* **32**, 1792-1797.
- El-Hajj, Z.W., Allcock, D., Tryfona, T., Lauro, F.M., Sawyer, L., Bartlett, D.H., and Ferguson, G.P. (2010). *Ann N Y Acad Sci* **1189**, 143-148.
- Emsley, P., and Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
- Evans, P. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
- Hanahan, D. (1985). *DNA cloning: A Practical Approach* IRL Press, Oxford, United Kingdom *vol. 1*, 109-135.
- Harris, J.B., LaRocque, R.C., Qadri, F., Ryan, E.T., and Calderwood, S.B. (2012). *Lancet* **379**, 2466-2476.
- Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., *et al.* (2000). *Nature* **406**, 477-483.
- Herbert, E.E., and Goodrich-Blair, H. (2007). *Nat Rev Microbiol* **5**, 634-646.
- Howard, A., O'Donoghue, M., Feeney, A., and Sleator, R.D. (2012). *Virulence* **3**, 243-250.
- Inoue, H., Nojima, H., and Okayama, H. (1990). *Gene* **96**, 23-28.
- Jacobsen, S.M., and Shirliff, M.E. (2011). *Virulence* **2**, 460-465.
- Kaasch, A.J., Dinter, J., Goesser, T., Plum, G., and Seifert, H. (2012). *Infection* **40**, 185-190.
- Kabsch, W. (1993). *J Appl Cryst* **26**, 795-800.
- Kim, S.E., Park, S.H., Park, H.B., Park, K.H., Kim, S.H., Jung, S.I., Shin, J.H., Jang, H.C., and Kang, S.J. (2012). *Chonnam Med J* **48**, 91-95.
- Komarova, N.V., Golubev, I.V., Khoronenkova, S.V., Chubar, T.A., and Tishkov, V.I. (2012). *Biochemistry (Mosc)* **77**, 1181-1189.
- Laemmli, U.K., and Favre, M. (1973). *J Mol Biol* **80**, 575-599.

- Lam, H., Oh, D.C., Cava, F., Takacs, C.N., Clardy, J., de Pedro, M.A., and Waldor, M.K. (2009). *Science* **325**, 1552-1555.
- Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., Reinhardt, R., Brinkhoff, T., Harder, J., Amann, R., *et al.* (2012). *ISME J* **6**, 2178-2187.
- Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 125-132.
- Murakami, K. (2012). *J Gastroenterol* **47**, 724-725.
- Nelson, M.D., and Fitch, D.H. (2011). *Methods Mol Biol* **772**, 459-470.
- O'Hara, C.M., Brenner, F.W., and Miller, J.M. (2000). *Clin Microbiol Rev* **13**, 534-546.
- Pepperell, C., Kus, J.V., Gardam, M.A., Humar, A., and Burrows, L.L. (2002). *Antimicrob Agents Chemother* **46**, 3555-3560.
- Pinhassi, J., Pujalte, M.J., Macian, M.C., Lekunberri, I., Gonzalez, J.M., Pedros-Alio, C., and Arahal, D.R. (2007). *Int J Syst Evol Microbiol* **57**, 2370-2375.
- Romer Villumsen, K., Dalsgaard, I., Holten-Andersen, L., and Raida, M.K. (2012). *PLoS One* **7**, e46733.
- Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J., and Studier, F.W. (1987). *Gene* **56**, 125-135.
- Rozalski, A., Kwil, I., Torzewska, A., Baranowska, M., and Staczek, P. (2007). *Postepy Hig Med Dosw (Online)* **61**, 204-219.
- Seifert, H., Strate, A., Schulze, A., and Pulverer, G. (1993). *Clin Infect Dis* **17**, 632-636.
- Su, Y.C., and Liu, C. (2007). *Food Microbiol* **24**, 549-558.
- Sugar, D.R., Murfin, K.E., Chaston, J.M., Andersen, A.W., Richards, G.R., deLeon, L., Baum, J.A., Clinton, W.P., Forst, S., Goldman, B.S., *et al.* (2012) *Environ Microbiol* **14**, 924-939.
- Timoney, P.J. (1996). *Comp Immunol Microbiol Infect Dis* **19**, 199-204.
- Williamson, E.D., and Oyston, P.C. (2012). *J Med Microbiol* **61**, 911-918.
- Zhao, J.S., Manno, D., Leggiadro, C., O'Neil, D., and Hawari, J. (2006). *Int J Syst Evol Microbiol* **56**, 205-212.

#### 4.3.1. DISCUSIÓN

En este trabajo se han realizado estudios cristalográficos, bioquímicos y bioinformáticos de dos racemasas de amplio espectro (BsrV y Bsr<sub>Ah</sub>) y se han identificado las bases estructurales que permiten a estas enzimas transformar un amplio rango de sustratos. También se ha determinado la huella molecular que permite identificar racemasas similares a BsrV de diversos microorganismos y distinguirlas de las alanina racemasas monoespecíficas.

Por otro lado, se ha descubierto que BsrV puede racemizar adicionalmente diez aminoácidos presentes en proteínas, así como varios aminoácidos no proteínogénicos. Los sustratos de BsrV se encuentran contenidos en dos clases principalmente: aminoácidos alifáticos no ramificados en  $\beta$  (Ala, Ser, Leu, Cys, Gln, Asn y Met) y aminoácidos básicos (Arg, Lys y His). El amplio rango de sustratos que acepta BsrV, incluyendo los no naturales (ornitina, norleucina, homoserina, *N*-acetil lisina metil éster, diaminobutirato y aminonutirato), en comparación con otras racemasas hace que esta racemasa se presente como objeto de aplicaciones biotecnológicas e industriales. Actualmente, la producción de DAA es un proceso caro que depende de una catálisis química ineficiente.

Curiosamente los miembros de esta familia de racemasas de amplio espectro se encuentra contenido dentro de las bacterias Gram-negativas. Sin embargo, algunos organismos Gram-positivos, como *B. subtilis*, se han propuesto como productores de NCDAA, este resultado sugiere que podrían existir otras familias de enzimas de amplio espectro aún no identificadas. El espectro de NCDAA que produce *B. subtilis* incluye varios aminoácidos que no pueden ser convertidos por la BsrV (D-Tyr y D-Trp).

Un análisis exhaustivo de las estructuras de las racemasas de amplio espectro, BsrV y Bsr<sub>Ah</sub>, con respecto a las mono-específicas, Alr<sub>Ah</sub>, Alr<sub>Ec</sub> y Alr<sub>Bs</sub>, sugiere que la plasticidad del mecanismo catalítico de estas racemasas tiene tres determinantes esenciales: la disposición y características estructurales del sitio de entrada, la anchura del canal principal hacia el sitio activo y su entorno catalítico, el cual es capaz de contener y estabilizar sustratos de naturaleza química diversa. La amplitud del sitio de entrada y el canal de la BsrV permite facilitar interacciones con aminoácidos más grandes que la alanina, de hecho, análisis estructurales, bioquímicos y de modelado sugieren que el canal de entrada de la BsrV podría también acomodar cadenas peptídicas. El sitio catalítico de las racemasas similares a BsrV está menos constreñido en comparación con las alanina racemasas, aparentemente debido a diferencias en los componentes de las cavidades (Cl<sup>-</sup> y Pro391) y a la disposición de los dominios enzimáticos que interaccionan. En trabajos previos se ha demostrado una variabilidad en la posición relativa de los dominios C y N terminal atribuida a cambios en los aminoácidos de la interfase del dímero. En las racemasas de amplio espectro, esta interfase está altamente alterada por la presencia de una nueva estructura formada por la extensión de la secuencia del N terminal. La maquinaria catalítica de las racemasas de amplio espectro presenta también diferencias importantes respecto a las enzimas monoespecíficas, como son la ausencia de la lisina



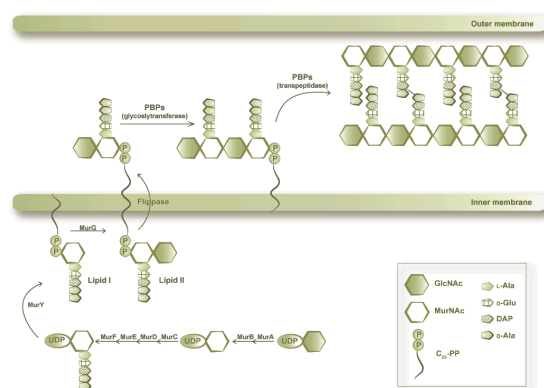
carbamilada, o la presencia de un ión  $\text{Cl}^-$  cercano al PLP. Análisis mutacionales indicaron que la extensión N terminal, así como la huella de 16 aminoácidos que permite identificar las enzimas de amplio espectro, son críticas para la actividad de la racemasa.

Dado el impacto de los NCDAA en una gran variedad de procesos celulares (estabilidad del biofilm, esporulación y comunicación celular, etc.), la caracterización estructural y regulatoria de las racemasas de amplio espectro podría abrir una puerta para el diseño de nuevas drogas que perturbarían la adaptación de la bacteria a condiciones adversas.

Las conclusiones de este trabajo son las siguientes:

- ✓ Se ha resuelto la estructura cristalográfica de la proteína BsrV, con unas características únicas que le permiten convertir una gran variedad de aminoácidos.
- ✓ Se ha resuelto la estructura cristalográfica de la racemasa de amplio espectro Bsr<sub>Ah</sub>.
- ✓ Se han establecido las bases estructurales que determinan una racemasa de amplio espectro. Además, se ha descrito la huella molecular que permitiría identificar enzimas de amplio espectro anteriormente anotadas como monoespecíficas.
- ✓ Se han identificado los determinantes esenciales que permiten la plasticidad del mecanismo catalítico de este tipo de racemasas, relativos a la disposición del sitio de entrada, la amplitud del canal y el entorno catalítico.

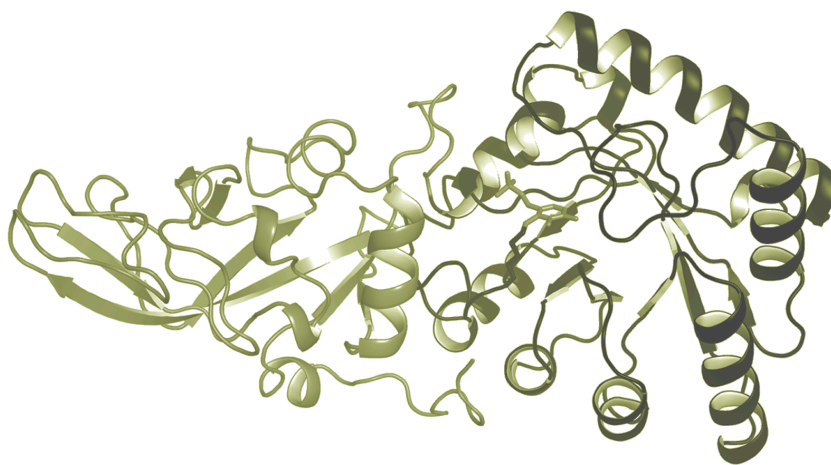
## Trabajo de Investigación IV:



## Structural bioinformatics in broad-spectrum racemases: a new path in antimicrobial research

Noelia Bernardo-García, Pedro Sánchez-Murcia, Federico Gago, Felipe Cava and Juan A. Hermoso

*Current Organic Chemistry* (2015). Accepted





#### 4.4. INTRODUCCIÓN A LA BIOINFORMÁTICA DE RACEMASAS DE AMPLIO ESPECTRO Y APORTACIÓN DEL AUTOR

Los D-aminoácidos son componentes esenciales de la pared celular, juegan un papel importante regulador en la esporulación, formación de biofilm y comunicación celular. Las racemasas son las enzimas que catalizan la conversión de los L-aminoácidos en D-aminoácidos. La mayoría de estas enzimas son monoespecíficas; sin embargo, existe una familia de racemasas de amplio espectro que son capaces de racemizar 10 de los 19 aminoácidos que presentan quiralidad. Estas racemasas pueden convertir una variedad de residuos diferentes, como son los aminoácidos alifáticos y cargados positivamente, produciendo lo que se conoce como D-aminoácidos no canónicos. Las estructuras cristalográficas junto con estudios bioinformáticos han permitido la identificación de residuos que definen la huella molecular de las racemasas de amplio espectro. En este trabajo se resumen los últimos conocimientos sobre esta familia en comparación con las ampliamente estudiadas alanina racemasas y se hace un análisis bioinformático de la cavidad catalítica y de su capacidad para racemizar sustratos muy diferentes.

A través de este trabajo se cumplieron los siguientes objetivos:

- ✓ Se ha realizado una revisión de los tipos de racemasas y su clasificación en función del mecanismo de reacción.
- ✓ Se han detallado los mecanismos catalíticos de ambos tipos de racemasas, mono específicas y multiespecíficas.
- ✓ Se ha realizado un análisis comparativo de las estructuras de la nueva familia de racemasas de amplio espectro con las ampliamente estudiadas alanina racemasas, detallando las diferencias estructurales que las definen.
- ✓ Se ha llevado a cabo un análisis bioinformático que avala los resultados obtenidos a través de los diferentes *docking* realizados con los diversos sustratos que acepta la racemasa de *V. cholerae* BsrV.

En esta publicación, la autora de esta Tesis ha realizado el análisis estructural, la caracterización estructural y funcional de las racemasas y la discusión de los resultados bioinformáticos.



**Structural bioinformatics in broad-spectrum racemases: a new path in antimicrobial research.**

Noelia Bernardo-García, Pedro Sánchez-Murcia, Federico Gago, Felipe Cava y Juan A. Hermoso

Current Organic Chemistry (2015) Aceptado.



## **Structural bioinformatics in broad-spectrum racemases: a new path in antimicrobial research**

Noelia Bernardo-García<sup>1</sup>, Pedro Sánchez-Murcia<sup>2</sup>, Federico Gago<sup>2</sup>, Felipe Cava<sup>3</sup> and Juan A. Hermoso\*<sup>1</sup>

*<sup>1</sup>Department of Crystallography and Structural Biology, Instituto de Química Física “Rocasolano”, CSIC, Madrid, Spain; <sup>2</sup>Área de Farmacología, Departamento de Ciencias Biomédicas, Unidad Asociada de I+D+I al CSIC, Universidad de Alcalá, Madrid, Spain; <sup>3</sup>Department of Molecular Biology and Laboratory for Molecular Infection Medicine Sweden, Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden*

\*Address correspondence to this author at the Department of Crystallography and Structural Biology, Instituto de Química Física “Rocasolano”, CSIC, c/ Serrano 119, E-28006 Madrid, Spain; Tel/Fax: +34 917 459 538/+34 915 642 431; E-mail: xjuan@iqfr.csic.es

**Running head:** Bioinformatics in broad-spectrum racemases

**Keywords:** Non-canonical D-amino acids; racemases; bioinformatics; antibiotics resistance; X-ray crystallography.



**ABSTRACT**

D-amino acids are essential components of the bacterial cell wall and play notable roles in microbiology as regulators, for example in sporulation, biofilm formation or interspecies communication. Racemases are the specific enzymes catalyzing the interconversion of L-amino acids to D-amino acids. While most of racemases are mono-specific, a family of broad-spectrum racemases that can racemize ten of the 19 natural chiral amino acids has been recently reported. These enzymes can interconvert radically different residues such as aliphatic and positively charged residues producing non-canonical D-amino acids. Crystal structures together with bioinformatics allowed identification of the residues defining the molecular footprint in broad-spectrum racemases, the specific features of their active sites and the structural basis of their promiscuity. Here we review the recent knowledge on this family compared with the well established of alanine racemases. This structural information is a prerequisite for the development of novel drugs against the important human pathogens for which broad-spectrum racemases play a key role.

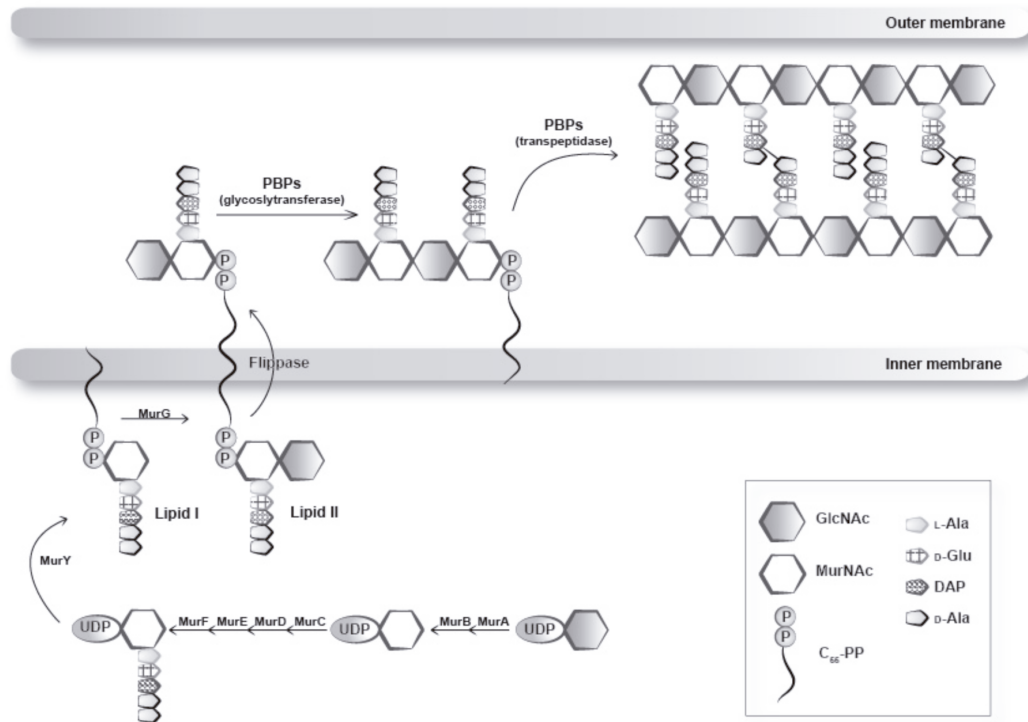
**Relevance of D-Amino Acids**

In nature, we normally find L-forms in amino acids and D-forms in sugars. Whereas, both enantiomers can be produced by chemical biosynthesis, there is substrate stereospecificity in many enzymes since they act on only one of the two enantiomers. L-amino acids are the basic units of proteins. Although ribosomal translation of proteins is fundamentally restricted to the use of L-amino acids, the D-forms are also widely present in microbes, plants, and animals.

D-amino acids are mirror images of L-amino acids, in which the chirality at the alpha carbon has been inverted. D-amino acids are highly resistant to protease-mediated degradation and have low immunogenicity. These findings and the fact that D-amino acids have recently emerged as extracellular effectors with remarkable roles in microbiology (e.g. cell wall regulators, sporulation, biofilm, interspecies communication...) make them attractive candidates for peptidomimetics drug design.

In nature, the majority of D-amino acids (DAA) are found in bacterial cell walls, where they are key constituents of the short peptides that crosslink the glycan chains within the peptidoglycan (PG) polymer (Vollmer et al., 2008). PG consists of repeating linear polymers of alternating  $\beta$ -linked *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) units crosslinked by short peptide stems. Depending on the species, the amino acid composition and type of crosslink varies (e.g. L-Ala- $\gamma$ -D-Glu-L-Lys(Gly)<sub>5</sub>-D-Ala-D-Ala in *S. aureus*; with the pentaglycyl moiety appended to the side chain of L-Lys) (Soda et al., 2001). PG biosynthesis starts in the cytoplasm, where the UDP-MurNAc-pentapeptide and UDP-GlcNAc are synthesized. Bactoprenol-linked UDP-disaccharide pentapeptide (lipid II) is formed to flip PG precursors across the membrane. Third, PG precursors are polymerized by the transglycosylases and transpeptidase activities of the synthetic penicillin-binding proteins (PBPs) resulting in

the murein sacculus (Fig. 1).



**Figure 1.** PG biosynthesis in Gram negative bacteria. PG precursors are synthesized in the cytoplasm; MurA, UDP-GlcNAc enolpyruvyl transferase and MurB, UDP-MurNAc dehydrogenase, synthesized UDP-MurNAc from UDP-GlcNAc. Subsequently, the Mur enzymes (MurC-F) are responsible of the Lipid I configuration. Once the Lipid II is formed a flippase translocates it to the periplasmic space; herein, the PBPs continue building the PG structure. Abbreviations: GlcNAc, *N*-Acetylglucosamine; MurNAc, *N*-Acetylmuramic acid; C<sub>55</sub>-PP, undecaprenyldiphosphate or bactoprenol; L-Ala, L-Alanine; D-Glu, D-Glutamic acid; DAP, diaminopimelic acid; D-Ala, D-Alanine.

In addition to “canonical” DAA such as D-Ala and D-Glu, which are present in virtually all bacteria, a number of species are known to produce cocktails of unusual DAA named noncanonical DAA (NCDAAs) that carry out a variety of regulatory functions. In *V.*

*cholerae*, NCDAAs edit PG architecture to control PG amount and strength (Lam *et al.*, 2009) during stationary phase. PG remodeling by NCDAAs occurs through the coordinated action of multispecific enzymes where the catalytic promiscuity of periplasmic transpeptidases and cytosolic ligases is key (Hashimoto *et al.*, 1993). In addition to the cell wall, DAAs are directly involved in important biological processes taking place in bacteria, such as bacterial growth, spore germination and biofilm formation. Virtually all bacteria form biofilms to cope with environmental adversities; DAAs can break up the biofilm previously arranged biofilm (Kolodkin-Gal *et al.*, 2010, Yang *et al.*, 2015). D-Ala has been reported to inhibit spore germination (Fey *et al.*, 1964). In addition, certain pathogens are surrounded by a capsule that protects them from phagocytosis thereby increasing their virulence. *Bacillus anthracis*' capsule is fully composed of an infrequent polymer of  $\gamma$ -linked D-glutamic acid ( $\gamma$ DPGA) which is produced after spore germination (Hubbard *et al.*, 2013).

D-amino acids are also reported to play roles in signaling. In fact, high concentration of D-serine and D-aspartate are found in mammals and humans in the central nervous system (Dunlop *et al.*, 1986, Fisher *et al.*, 1991, Grishin *et al.*, 1995). D-serine metabolism is associated with different disorders, such as schizophrenia, ischemia, epilepsy, and other neurodegenerative diseases (Fuchs *et al.*, 2005). Also, the accumulation of abnormal proteins containing D-amino acids in their composition is associated with age-related ocular disease provoking decreased transparency (Kaji *et al.*, 2010).

Alternatively, many enzymes related to D-amino acids biosynthesis such as D-hydantoinase, D-amino acid amidohydrolase, D-peptidase, D-amino transferase and others are finding wide applications in industry (Gao *et al.*, 2015). The production of

optically active amino acids is required by soem pharmaceutical companies for the synthesis of drugs, pharmaceuticals, food additives and drug intermediates. Microbial fermentation rather than chemical synthesis has been used to produce amino acids, especially in large scale, and different L-amino acids such as lysine, threonine, glutamic acid, tryptophan, and phenylalanine (Hemraj S. Nandanwar, 2005).

Of special relevance is the synthesis of non-ribosomal peptides (NRPs) by the NRP synthases. NRP synthesis is an alternative pathway that allows production of polypeptides by means other than through the traditional translation mechanism. NRPs are composed of cyclic and/or branched structures, can contain D-amino acids, present *N*-methyl and *N*-formyl groups, or can be glycosylated, acylated, halogenated or hydroxylated. Non-ribosomal peptides are a very diverse family of natural products with an extremely broad range of biological activities and pharmacological properties. Numerous NRPs are used commercially as antibiotics (actinomycin D), cytostatics, and immunosuppressants. Soil bacteria actively produce NRPs actively, as exemplified by *Serratia plymuthica*, which is able to kill the nematode *Caenorhabditis elegans* through production of secondary metabolites such as zeamine (Hellberg *et al.*, 2015).

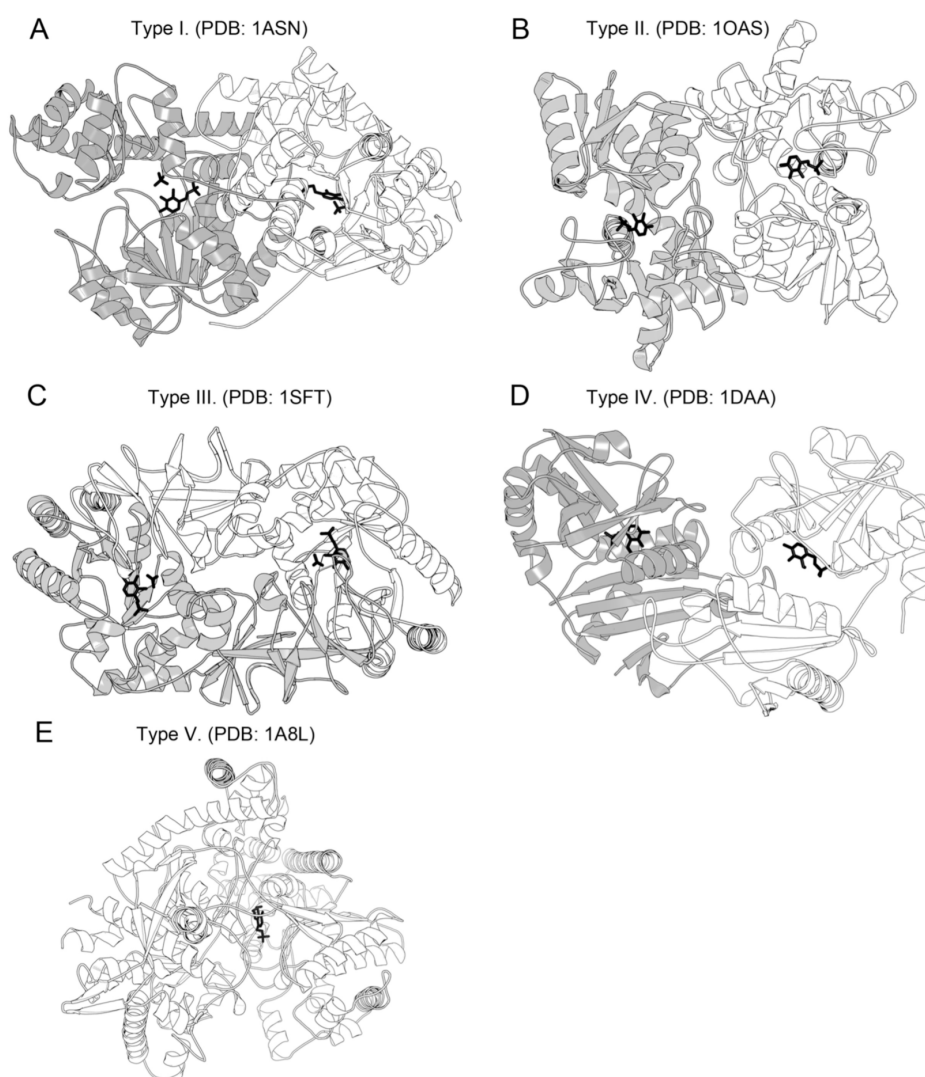
### **Racemases definition and classes**

Racemases are the specific enzymes that catalyze the interconversion between L-amino acids and D-amino acids. These enzymes can be classified into two main classes depending on whether their mode of action is dependent or independent of coenzyme pyridoxal 5'-phosphate (the biologically active form of vitamin B<sub>6</sub>). The pyridoxal 5'-phosphate (PLP) dependent racemases group together eukaryotic enzymes such as aspartate racemase (AspR), alanine racemase (AlaR) and serine racemase (SerR), and also the bacterial members such as AlaR, SerR, ArgR and the recently identified new class of broad-spectrum racemases (BsR). Among the PLP-independent enzymes are

aspartate racemases (AspR), glutamate racemases (GluR) and proline racemases (ProR). Both PLP-dependent and PLP-independent types of enzymes deprotonate the  $\alpha$ -carbon of the L-amino acid to reprotonate it on the opposite side, thus generating the D-form of the amino acid.

### PLP-dependent racemases. Types and fold.

Considering their three-dimensional structures the PLP-dependent racemases have been



**Figure 2.** Three-dimensional structures from PLP-dependent enzymes, folding types I-V. The structures of PLP enzymes would correlate with its reaction type. PLP is depicted in black sticks.

classified into five types (fold type I-V) (Grishin *et al.*, 1995, Soda *et al.*, 2001). Type I appears in aminotransferases and decarboxylases and is represented by aspartate aminotransferase (Fig. 2A). The biological active unit is a homodimer, each monomer presenting a large domain composed of a seven-stranded  $\beta$ -sheet and a small domain with a three-to-four-stranded  $\beta$ -sheet decorated with  $\alpha$ -helices on one side. The active site, located at the interface between monomers, always contains an aspartate residue that interacts with the pyridine nitrogen of PLP.

The type II is represented by tryptophan synthase (Fig. 2B) and also contains serine racemase. The biological active unit is a homodimer but, differently from type I, each active site is completely formed by residues from a single monomer. Another difference with the active site of type I enzymes is that the pyridine nitrogen of PLP makes a H-bond with a serine residue.

The type III contains AlaR, ArgR, some amino acid decarboxylases and the new family of broad-spectrum racemases (Fig. 2C). AlaR is the prototype for this group and its biological unit is a homodimer. Each monomer is composed of two domains, one of them folded into an eight-stranded  $\alpha/\beta$  barrel, and the other made up of  $\beta$  strands. The active site in the dimer is built with residues from both monomers.

The biological unit of type IV enzymes is also formed by a homodimer (Fig. 2D). Each monomer presents two domains; one of them folded into an antiparallel six-stranded  $\beta$  sheet and the other containing a pseudo  $\beta$  barrel surrounded by  $\alpha$  helices. Active sites are completely arranged in each monomer and are located at the interface between the N- and C-terminal domains. D-Ala aminotransferases, among others, belong to this group.

Type V presents a large monomer as a single biological unit (Fig. 2E). Contrary to the other four types, the catalytic role of PLP is played just by its phosphate moiety. Glycogen phosphorylase is the prototype for this group of enzymes.

### **The catalytic mechanism of alanine racemases**

The first crystal structure of a PLP-dependent amino acid racemase, AlaR<sub>Bs</sub> from *Geobacillus stearothermophilus* (formerly known as *Bacillus stearothermophilus*), was solved in 1997 in its holo form (Shaw *et al.*, 1997b, Shaw *et al.*, 1997a) and suggested that this dimeric enzyme uses two amino acid residues, Lys39 from one monomer and Tyr265' from the other monomer, for catalysis (Scheme 1). While the N $\epsilon$  of Lys39 was shown to be the point of covalent attachment to PLP, via an aldimine linkage, the role of Tyr265' was postulated to be a 2-proton donor to the carbanion intermediate. The mechanism initially proposed for AlaR involved the formation of an *N*-protonated quinonoid intermediate, in analogy to previous proposals for other PLP-dependent enzymes such as aminotransferases, despite the fact that Arg219, an arginine that is highly conserved in fold type III racemases, was observed at hydrogen-bonding distance from the pyridine nitrogen of the cofactor. In fact, the quinonoid intermediate could not be experimentally detected in either the wild-type AlaR<sub>Bs</sub> or the Arg219Gln and Arg219Ala variants but its presence was actually demonstrated in a Arg219Glu variant (Sun & Toney, 1999, Watanabe *et al.*, 1999, Watababe *et al.*, 1999).

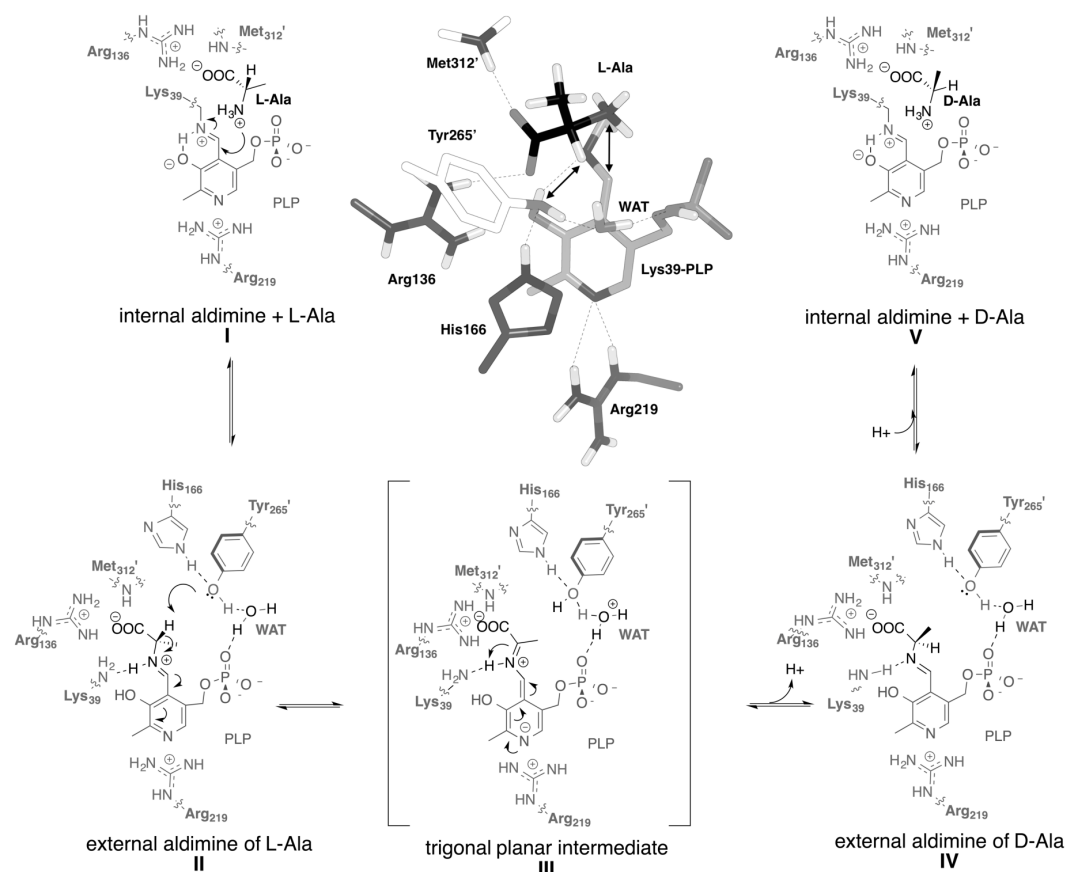
A few years later, on the basis of two co-crystal structures of AlaR<sub>Bs</sub> with bound external aldimines formed between PLP and either L-Ala or D-Ala (PDB codes 1L6F and 1L6G, respectively) and molecular orbital calculations, a revised mechanistic proposal was made (Watanabe *et al.*, 2002b) that nullified the need to invoke a quinonoid intermediate but maintained the participation of Lys39 and Tyr265' as catalytic bases for  $\alpha$ -proton abstraction. In this model, the carboxylate of the substrate



was assigned the role of participating in catalysis directly by mediating the proton transfer between these two residues (i.e. abstraction from the phenolic oxygen of Tyr265' and donation to the amino group of Lys39). Nonetheless, in these crystal structures the Tyr265' phenol is clearly seen accepting a hydrogen bond from the imidazole of His166 and this interaction forces the OH hydrogen away from each amino acid carboxylate. On the other hand, the very short distances ( $< 3.0 \text{ \AA}$ ) between the Ne of Lys39 and the amino N of the bound amino acids (L-Ala or D-Ala) can be taken as an unequivocal indication that a proton exists between both atoms which rules out the simultaneous existence of an external aldiminium ( $\text{C}=\text{NH}^+$ ) and a fully protonated Lys39 ( $\text{NH}_3^+$ ) on electrostatic grounds.

This and additional structural work has been elegantly complemented over the years by extensive site-directed mutagenesis experiments, kinetic isotopic effects (Faraci & Walsh, 1988, 1989, Shaw *et al.*, 1997b, Stamper *et al.*, 1998, Sun & Toney, 1999, Watanabe *et al.*, 1999, Watababe *et al.*, 1999, Watanabe *et al.*, 2002b, Spies & Toney, 2003, Spies *et al.*, 2004, Spies & Toney, 2007, Fenn *et al.*, 2005), and also computational chemistry methods (Spies & Toney, 2007, Major *et al.*, 2006). Taking together all the evidence brings into question some of the earlier assumptions reported above and casts doubts on the proposed two-base mechanism. In fact, it is still unclear how the catalytic Tyr265' can accept the proton from the substrate amino acid. Moreover, the protonation states of the external aldimine and the catalytic lysine have generally been vaguely defined (e.g. the total number of protons in the catalytic cycle is not always conserved) and the presence of tightly bound water molecules in the active site has been ignored for the most part. For these reasons, we think it is in order to summarize in the following paragraph what appears to us as the most convincing scenario for catalysis using AlaR<sub>BS</sub> as the reference system.

By facilitating a substantial increase in the acidity of the  $\alpha$ -proton in L- and D-amino acids, AlaR's are able to interconvert one enantiomer into the other through a Brønsted acid/base mechanism involving the coenzyme pyridoxal phosphate (PLP) and the side chains of two conserved residues, namely a tyrosine and a lysine (Tyr265' and Lys39 in AlaR<sub>Bs</sub>), which are located on opposite faces of the PLP ring (central Figure in Scheme 1). In the substrate-free form, the aldehyde group of PLP forms an imine linkage (Schiff base) with the  $\epsilon$ -amino group of this catalytic lysine (internal aldimine **I**, Scheme 1). Upon substrate (e.g. L-Ala) binding, its amino group undergoes a transaldimination reaction (an imine exchange) to yield the external aldimine **II**. This substrate-coenzyme adduct is stabilized by a favorable hydrogen bond between the phenol of PLP and the imine nitrogen. The L-Ala carboxylate, on the other hand, is hydrogen-bonded to the guanidinium of Arg136 and the backbone nitrogen of Met312'. The  $\alpha$ -proton of L-Ala is then abstracted by the Tyr265' phenol, which is activated with the assistance of the imidazole from a neighbouring His (His166 in AlaR<sub>Bs</sub>) and a water molecule (WAT<sub>463</sub> in monomer A, PDB entry 1SFT, and WAT<sub>2300</sub> in BrsV, PDB entry 4BEU). This water is hydrogen-bonded to a phosphate oxygen of PLP as well. The negative charge on the ensuing carbanion (trigonal planar intermediate **III**) is delocalized not only to the carboxylate group of the substrate amino acid but, more importantly, to the expanded network of conjugated  $\pi$ -bonds in PLP. The reprotonation of **III** by the amino group of Lys39, located on the opposite side of the PLP ring, results in inversion of absolute configuration at the amino acid  $\alpha$ -carbon (external aldimine **IV**). Finally, a second transaldimination reconstitutes the native form of the enzyme (internal aldimine **V**) and gives rise to the D-Ala product, which is released from the active site.



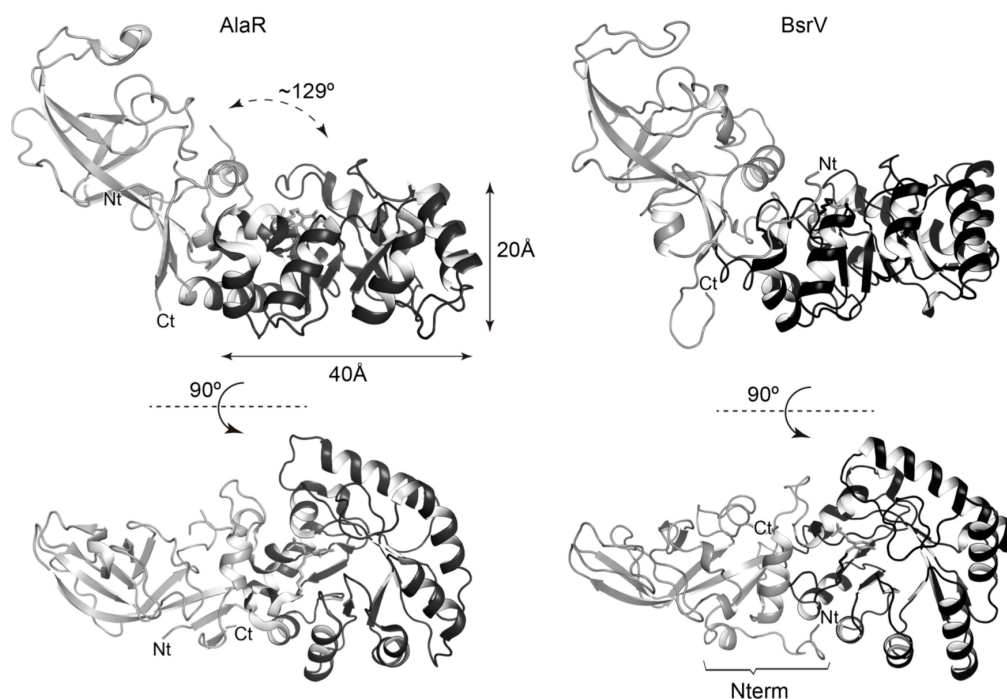
**Scheme 1.** Proposed reaction mechanism for the racemization of L-Ala by AlaR<sub>Bs</sub>. Dotted lines indicate hydrogen bonds and curved arrows depict the movement of electrons. To better understand the spatial relationships a detail of the active site in a 3D molecular model of complex **I** (top, middle) is also shown. The double-headed arrows in this figure denote the distances between the C $\alpha$  hydrogen and the Brønsted acid (Tyr265') and between the substrate amino group and the aldehyde carbon of PLP.

### Structural features of monofunctional alanine racemases

Alanine racemase (AlaR) belongs to the type III of PLP-dependent racemases. Bacterial AlaR enzymes have been deeply studied considering their relevance in production of D-Ala, one of the main components of the peptide-stem in the bacterial cell wall of both

Gram-negative and Gram-positive bacteria, and also as important targets for the discovery of new antibacterial drugs. Crystallographic structures for different bacterial AlaR, alone and in complexes with substrate analogues have been already reported. They provide high-resolution information about their fold, active site and catalytic machinery.

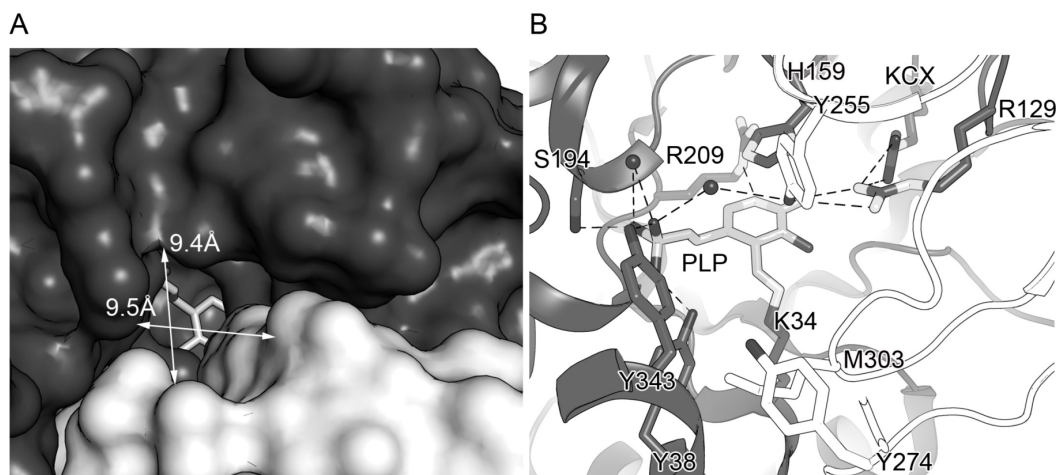
The enzyme acts as a homodimer, each monomer being composed of two different domains. The N-terminal domain is folded into an eight-stranded  $\alpha/\beta$  barrel that can be roughly described as a flat cylinder of  $\sim 40\text{\AA}$  diameter and  $\sim 20\text{\AA}$  height (Fig. 3). The C-terminal domain presents architecture of six-stranded  $\beta$  barrel plus two additional  $\beta$  strands (one from the N-terminus and the other from the C-terminus) that extend the surface of this domain. Both domains in the monomer make an angle of  $\sim 129^\circ$  (Fig. 3). The AlaR dimer is built in such a way that the N-terminal domain of a monomer is facing the C-terminal domain of the partner (Fig. 2C). The active site, located at the N-terminal domain, is partially covered by the C-terminal domain of the other monomer resulting in a tight cavity that allows the L-Ala substrate to reach the PLP located at the bottom of the cavity (Fig. 4A).



**Figure 3.** Structural comparison of AlaR and BsrV monomers. The N-terminal domain (colored in black) and the C-terminus (colored in white) are represented as cartoons. The N-terminal insertion (Nterm) present in the BsrV structure results in a closer arrangement of both domains in the dimer.

*The active site of AlaR.* The PLP cofactor is strongly stabilized in the active site cavity. A Lys residue (Lys34 in *Escherichia coli* AlaR, AlaR<sub>Ec</sub>) makes a covalent bond with PLP (Fig. 4B). The pyridine nitrogen of PLP makes a H-bond with an arginine residue (Arg209 in AlaR<sub>Ec</sub>), reinforcing the idea that the protonation state of the pyridine nitrogen of PLP does not take place during catalysis. The PLP ring is sandwiched between hydrophobic residues on one side and a histidine residue (His159 in AlaR<sub>Ec</sub>) on the other side (Fig. 4B). The phosphate moiety of PLP is stabilized by many polar interactions with protein residues (Tyr38, Tyr343 and Ser194) and with water molecules (Fig. 4B). An arginine residue (Arg129 in AlaR<sub>Ec</sub>) is oriented towards the PLP and plays an important role in stabilization of the carboxylate moiety of the L-Ala substrate in the quinonoid intermediate. In order to keep this arginine oriented in catalysis,

another peculiar residue is placed in the AlaR active site: an N-carboxylated lysine (KCX, Fig. 4B) that makes a double H-bond with Arg129. The Tyr255 residue (Fig. 4B) from the other monomer also plays an important role in stabilization of the quinonoid intermediate.



**Figure 4.** Active site in alanine racemases. (A) The molecular surface of the entry sites cavity in *Escherichia coli* AlaR (AlaR<sub>Ec</sub>) is depicted in black for one monomer and in white for the dimeric partner. The PLP cofactor is depicted as sticks. Dimensions of the entry cavity are labeled. (B) Details of the active site of AlaR<sub>Ec</sub>. Relevant residues involved in the catalytic mechanism and the PLP cofactor are displayed as sticks, and water molecules are shown as black spheres.

#### Inhibitors in mono-specific racemases

Small molecules such as cycloserine and O-carbamyl-D-serine are the two natural antibiotics known to inhibit alanine racemase. The first one has been developed commercially to treat tuberculosis. Cycloserine is not target-specific, and some efforts have been made to improve its activity or specificity through alteration of the side chain attached to the cycloserine ring. Other alanine analogue inhibitors are  $\beta,\beta,\beta$ -trifluoroalanine, alanine phosphonate, L-amino-cyclopropane phosphonate and  $\beta$ -chloro- and  $\beta$ -fluoroalanine. Recently, some non-substrate alanine racemase inhibitors

against *Mycobacterium tuberculosis* have been identified (Anthony *et al.*, 2011). These novel inhibitors include hydrazide, hydroxamic acid, thiopyridine, pyrimidine carboxamide, thioamide, pyridine ester, oxadiazole N-oxide, carbamodithioate, isothiazolopyridione, benzopyranone oxime, alkoxy phenol, and quinolinedione; seven of them inhibit *M. tuberculosis* growth and some are minimally cytotoxic against mammalian cells (Anthony *et al.*, 2011).

Besides, small aliphatic hydroxamic acids, as well as L-aspartic acid  $\beta$ -hydroxamate, a competitive and discriminating inhibitor, have been reported to act as potent serine racemase inhibitors (Hoffman *et al.*, 2009). The mammalian serine racemase produces D-serine, a neurotransmitter that activates N-methyl-D-aspartate (NMDA) receptors in the central nervous system (Hoffman *et al.*, 2009). Since numerous neuropathologies have been related to an abnormal regulation of the NMDA receptors, serine racemase inhibitors would be a valuable instrument to investigate their operating mechanism and find a treatment for these pathologies.

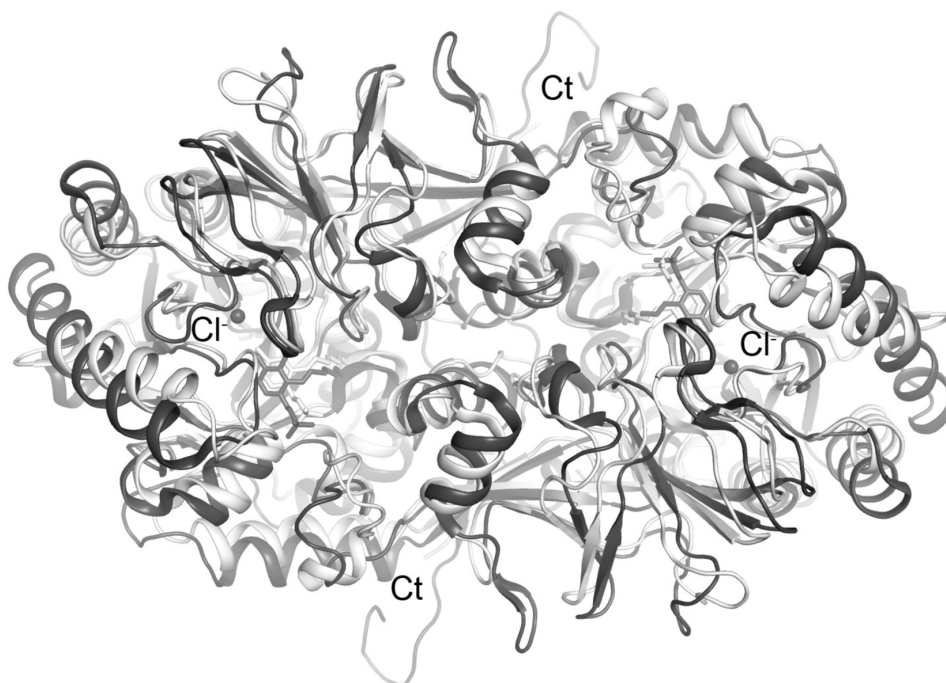
In the case of *Trypanosoma cruzi*, the causative agent of Chagas disease, a proline racemase play a key role in infectiveness and differentiation of the parasite through interactions with the mammalian cells in vitro. Pyrrole-2-carboxylic acid is a specific proline racemase inhibitor, which acts by decreasing the ability of *T. cruzi* to invade host cells (Coutinho *et al.*, 2009).

### **The new family of broad-spectrum racemases**

A new family of racemases has been recently described presenting large substrate promiscuity. The first family member is the broad-spectrum racemase from *Vibrio cholerae* (BsrV). Unlike AlaR, which only catalyzes the conversion of L-Ala to D-Ala,

BsrV can racemize ten amino acids (Ala, Leu, Met, Ser, Cys, Gln, Asn, His, Lys and Arg) of the 19 natural chiral amino acids (Espaillat *et al.*, 2014). Interestingly, BsrV can interconvert radically different residues such as aliphatic (the non- $\beta$ -branched ones) and positively charged residues. However, negatively charged, aromatic or  $\beta$ -branched aliphatic (Ile, Val and Thr) amino acids are not substrates of BsrV.

The 1.15 Å-resolution crystal structure of BsrV revealed the specific structural features defining broad-spectrum racemases (Espaillat *et al.*, 2014). The overall fold of BsrV is quite similar to that of type III mono-specific enzymes (closest structural homologue of BsrV is AlaR from *Bacillus anthracis* showing a root-mean-square deviation of 2.0 Å for 361 C alpha atoms). As other members of type III family, BsrV presents a dimeric arrangement in which the N-terminal domain of one monomer interacts with the C-terminal domain of the other monomer to build the active site. However, relevant differences are observed in the oligomeric arrangement when BsrV and AlaR are compared (Fig. 5).

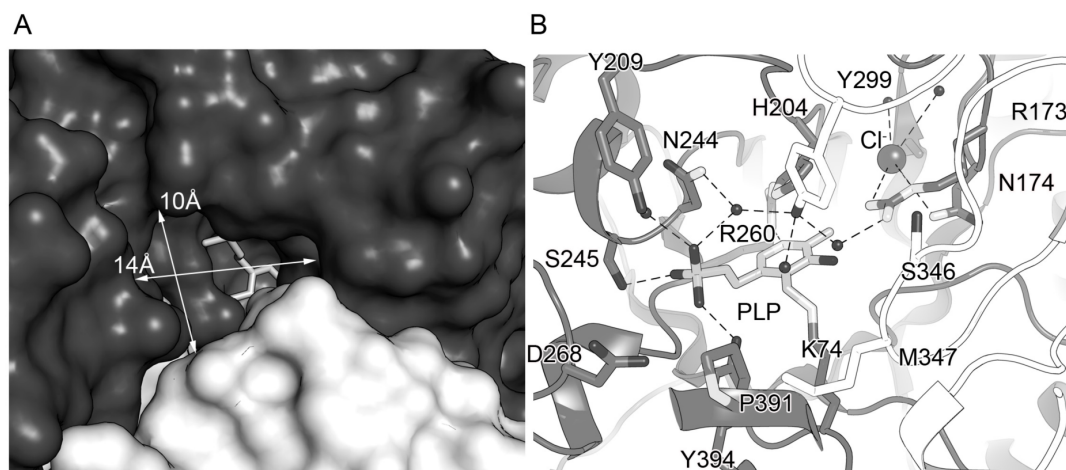




**Figure 5.** Structural superimposition of BsrV (black cartoon) and AlaR<sub>Ec</sub> (white cartoon). PLP is depicted as sticks and Cl<sup>-</sup> ions as spheres.

The BsrV monomer presents similar N- and C-terminal domains to those found in AlaR (Fig. 3). Besides, BsrV presents an extra  $\alpha$ -helix and a coiled-coil extension at the N-terminus (named Nterm in Fig. 3) that is important in the arrangement of the BsrV dimer. The amino acids in this extension act as a hinge-like structure that increases by 9° the relative arrangement of the N- and C-terminal domains observed in AlaR and produces an average displacement of 1.5 Å for the whole protein backbone (Espaillat *et al.*, 2014). The addition of this N-terminal structure in BsrV together with shorter loops in the side walls of the catalytic entry results in both a two-fold wider entry channel in BsrV versus that of AlaR (Fig. 6A) and also in a different separation between active sites in the dimer (37.6 Å vs. 45 Å in AlaR and BsrV, respectively).

*The active site of BsrV.* As in AlaR, the BsrV active site stabilizes the PLP cofactor by virtue of a covalent bond with a Lys residue (Lys74) (Fig. 6B) and is also sandwiched between hydrophobic residues on one side and by a His residue (His204) on the other. The pyridine nitrogen is also stabilized by a hydrogen bond with an arginine (Arg260) (Fig. 6B). However the phosphate group is differently stabilized in AlaR and BsrV. Only one Ser residue and one Tyr residue (Tyr394) interact with the phosphate (Figure 6B), while in BsrV a Pro residue (Pro391) replaces the second Tyr found in AlaR. Replacement of the bulky Tyr residue by a Pro residue enlarges the active site of BsrV, thus allowing substrates with bulky side chains. BsrV lacks the N-carboxylated lysine and the substrate-binding arginine residue (Arg173) is stabilized by a chloride ion (Fig. 6B). This Cl<sup>-</sup> ion is held in place by water molecules, two H-bonds with Arg173 and the side chain of Asn174. As will be explained below, these changes provide a great plasticity to the catalytic machinery of the broad-spectrum racemases.



**Figure 6.** Active site in broad-spectrum racemases. (A) Molecular surface of the entry cavity, one monomer colored in black and the other monomer in white. PLP is represented in sticks. Dimensions of the entry cavity are labeled. (B) Active site of BsrV with PLP cofactor and relevant residues involved in the catalytic mechanism depicted in sticks,  $\text{Cl}^-$  ion and water molecules are shown in gray and black spheres, respectively.

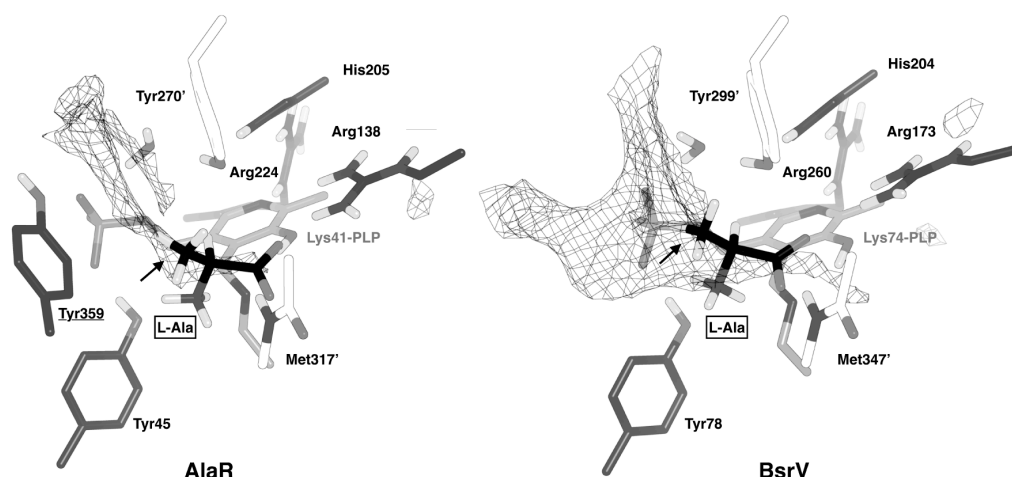
### Bioinformatic tools identify the molecular signature of broad-spectrum racemases

Bioinformatic analyses allowed the identification of the residues defining the molecular footprint in broad-spectrum racemases (Espaillat *et al.*, 2014) from a preliminary set of 1355 protein sequences describing a representative and non-redundant set of PLP-dependent racemase sequences. Multiple sequence alignment allowed clustering of these racemases in three subfamilies (Espaillat *et al.*, 2014). Subfamily 1, the most numerous, can be ascribed to the monospecific alanine racemases such as AlaR<sub>Ec</sub>, while Subfamily 2 would correspond to the BsrV-like proteins presenting a broader substrate spectrum. A new subfamily (Subfamily 3) was also identified without any of its members been biochemically characterized up to now. In the analysis a few number of outlier proteins (15) could not be allocated into any of the three subfamilies.

Comparative analyses between AlaR-like (subfamily 1) and BsrV-like (subfamily 2) members identified 16 sites making up the molecular footprint that allows distinction between both classes. Interestingly, among these 16 residues, two of them (positions 119 and 121) are involved in the oligomeric arrangement, 4 are directly involved in the active site entry and the remaining ten positions are engaged in catalysis. The identified residues could serve both to distinguish between alanine and broad-spectrum racemases and to design of specific ligands targeting broad-spectrum racemases. It is worth to mention that the BsrV-like family contains important Gram-negative pathogens such as the Vibrionaceae family, *Yersinia pseudotuberculosis*, *Yersinia pestis*, *Helicobacter hepaticus* and *Acinetobacter baumannii*, among others.

### **Bioinformatics explain differences in substrate selectivity between BsR and AlaR**

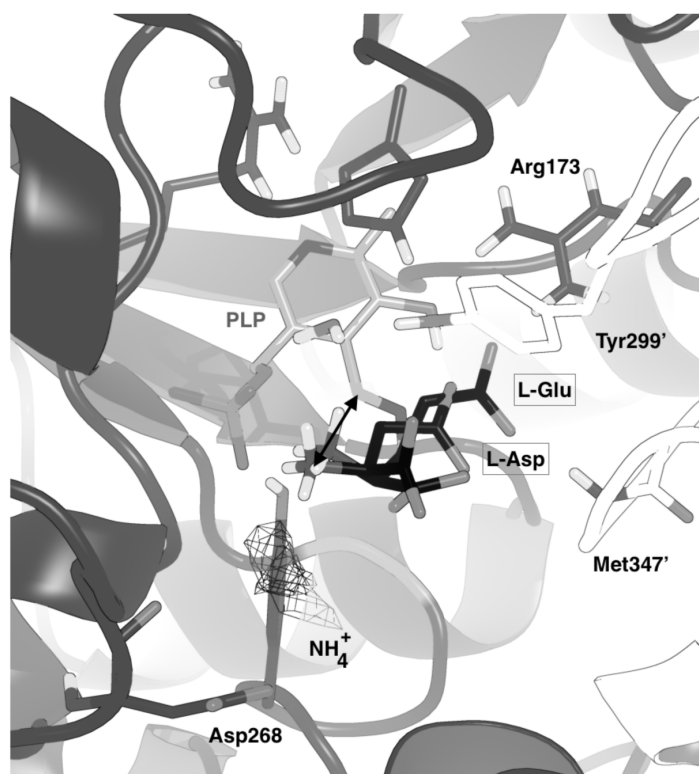
Designing of specific inhibitors, or drugs, targeting broad-spectrum racemases require understanding of the molecular differences in substrate selectivity between the mono-specific and the broad-spectrum enzymes. The combined use of high-resolution structures and bioinformatic tools can provide valuable information on this matter. Figure 7 shows the affinity maps, or molecular interaction fields, in the active site regions of AlaR<sub>Ba</sub> and BsrV calculated for a lipophilic probe (CH<sub>3</sub>). Each map occupies the space available for ligand binding above the phosphate moiety of PLP. Whereas in AlaR<sub>Ba</sub> this volume is confined between Tyr270' and Tyr359, it expands notably in BsrV because the equivalent position for the latter residue is occupied by a proline (Pro391).



**Figure 7.** Distinct active site architectures are responsible for the observed substrate selectivity differences between *B. anthracis* AlaR (PDB code 3HA1) and BsrV (PDB code 4BEU). The affinity maps calculated for a CH<sub>3</sub> probe using our in-house program cGRILL (Cortes-Cabrera *et al.*, 2015a), shown here as a wire mesh contoured at -2.0 kcal·mol<sup>-1</sup> in both figures, reveal the larger cavity available for amino acid substrates in the latter enzyme due to the smaller size of the proline residue compared to the tyrosine. Note the arrows pointing to the C $\beta$  atom of docked L-Ala (C atoms colored in black).

Regarding the substrate specificity of BsrV, the maps generated with the positively charged hydrogen-bonding donor probe help understand why Arg and Lys amino acids are accepted substrates for Bsr's (Wu *et al.*, 2012) whereas Asp and Glu are not and require a different class of non-PLP-dependent enzymes for racemization (Conti *et al.*, 2011, Spies *et al.*, 2004). These maps nicely match the highly negative molecular electrostatic potential generated by the PLP phosphate group and the side-chain carboxylate and carbonyl oxygen of Asp268 (Fig. 8) in the expanded volume available for the side chains of the substrates. When Asp and Glu are docked into the active site using program Glide (LLC, 2015) the best poses show that the standard binding mode is prevented due to electrostatic repulsion in this region and an alternate, non-productive

orientation is achieved instead (Fig. 8). In these binding poses, the side-chain carboxylates of Asp and Glu occupy positions that are equivalent to that of the  $\alpha$ -amino group of the other incoming amino acids (cf. Figs. 7 and 8).

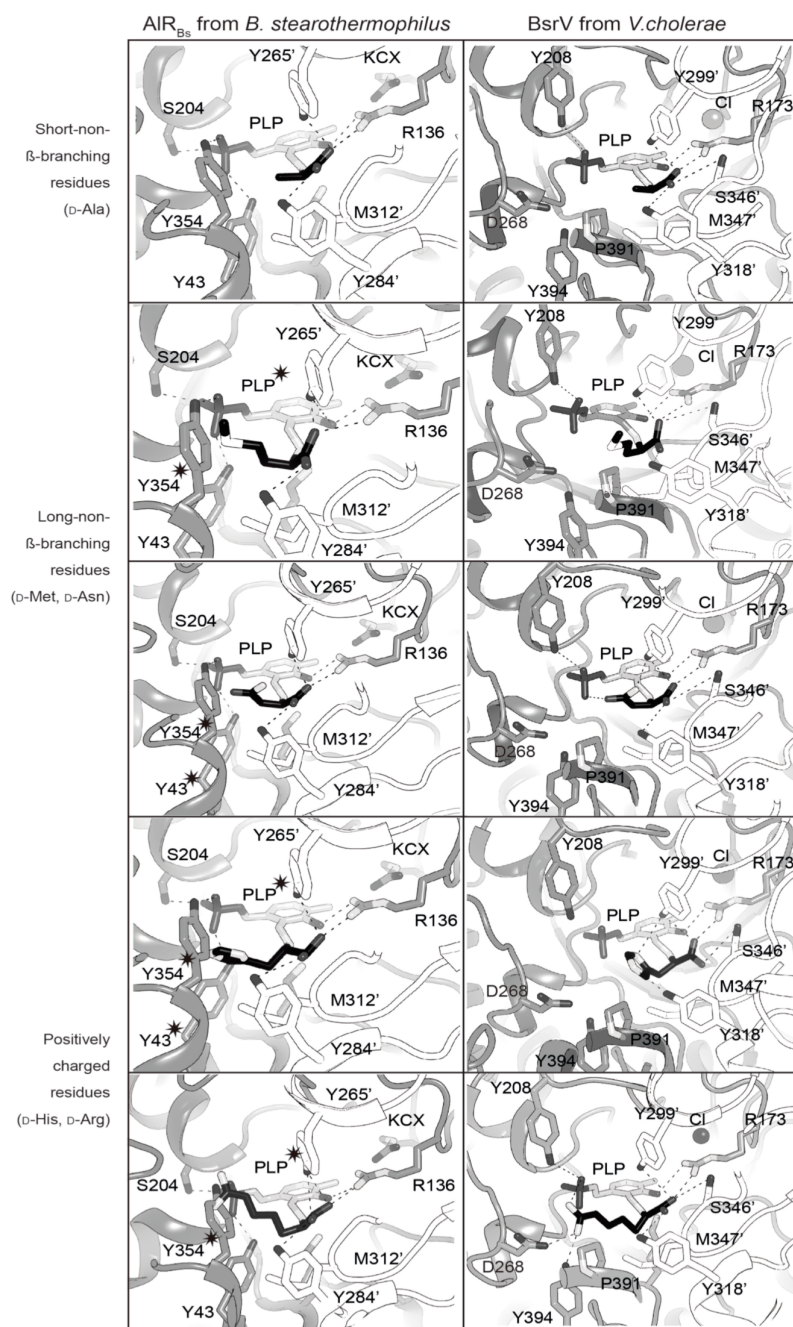


**Figure 8.** Detail of the best docking solutions for Asp and Glu (C atoms colored in black) in the active site of BsrV. Note that the binding poses are not compatible with a productive orientation of the  $C\alpha$  for catalysis (cf. Fig. 7) because the region where the side chain carboxylates would have to be located has a strong electronegative character, as highlighted by the complementary affinity map calculated with program cGRILL (Cortes-Cabrera *et al.*, 2015a, b) for a positively charged hydrogen bond donor ( $NH_4^+$  mesh). The double-headed arrows denote the distance between the  $\alpha$ -amino group of the incoming Asp or Glu and the aldehyde carbon of PLP.

In order to find out the structural features determining the substrate stabilization in mono-specific and multi-specific racemases, docking calculations were done by exploiting the alanine racemase structure from *Bacillus stearothermophilus* (AlaR<sub>Bs</sub>) in complex with D-alanine (Watanabe *et al.*, 2002a). BsrV was superimposed on this complex replacing substrate by all possible kind of amino acids (Fig. 9). For comparison purposes the same docking calculations were performed with AlaR (Fig. 9). Whereas AlaRs only interconverts alanine and serine, BsrV could fit ten of the 19 natural amino acids (Espaillat *et al.*, 2014), which include short-non- $\beta$ -branching residues (Ala, Ser, Cys, Leu), long-non- $\beta$ -branching residues (Met, Asn, Gln) and positively charged residues (His, Arg and Lys).

Docking analysis further explains the mechanism by which a broad range of amino acids is accommodated in the active site of BsrV. The carboxylate moiety of the substrate is stabilized by polar interactions with Arg173, Tyr299' and Tyr318' (Fig. 9). In aliphatic amino acids (Ala, Ser, Cys, Leu, Gln and Met), the side chains are stabilized by hydrophobic interactions with Met347' (Fig. 9). Basic amino acids (Arg, Lys and Orn) are further stabilized by a hydrogen bond with Tyr394 and by electrostatic interactions with Asp268 and the PLP phosphate group (Fig. 9). This matching is in agreement with the affinity maps calculated by cGRILL (Fig. 8).

Moreover, the docking results also provide a rationale to the finding that BsrV, but not Alrs, can accommodate a variety of amino acids (containing both long aliphatic and basic side chains). Tyr354 residue in AlaR<sub>Bs</sub> is replaced by a Pro residue (Pro391) in BsrV (Fig. 9). The short side-chain of Pro391 in BsrV (Ala residue in other BsRs) provides an enlargement of the active site and permits the correct fitting for substrates with different side-chains while the same substrates in AlaRs result in many clashes with residues building the active site cavity (Fig. 9).



**Figure 9.** DAA in BsrV *versus* AlaR<sub>Bs</sub>, docking experiments. Monomers are depicted in grey and white; each substrate is colored in black. Residues involved in substrate recognition are labeled and interactions are highlighted as dotted lines. Asterisks indicate substrate clashes.

**Conclusions**

The relevance of D-amino acids in bacterial survival as well as in several human diseases promoted the study of the structure and mechanism of the enzymes involved in amino acid racemization. Nowadays a large body of knowledge is available about mono-specific racemases, and especially on AlaR and its inhibitors. Recently a new family of periplasmic racemases has been described in a number of important human pathogens. Contrary to the mono-specific enzymes, the broad-spectrum racemases can convert ten amino acids to produce NCDAAs. High-resolution structures already reported together with knowledge obtained from bioinformatics provide important clues about substrate binding and catalysis in these promiscuous enzymes. Considering the relevant role of NCDAAs in cell-wall regulation and bacterial metabolism, this new information is paving the way to the design of novel and more effective antibacterials.

**CONFLICT OF INTEREST**

The author(s) confirm that this article content has no conflicts of interest.

**ACKNOWLEDGMENTS**

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness (BFU2011-25326) to J.A.H. and from the Biomedicine Program of the Autonomous Community of Madrid (S2010/BMD-2457) to J.A.H. and to F.G.. F. Cava is supported by the Laboratory for Infection Medicine Sweden (MIMS), by the Knut and Alice Wallenberg Foundation (KAW), and the Swedish Research Council.

**REFERENCES**



- Anthony, K. G., Strych, U., Yeung, K. R., Shoen, C. S., Perez, O., Krause, K. L., Cynamon, M. H., Aristoff, P. A. & Koski, R. A. (2011). *PloS one* **6**, e20374.
- Conti, P., Tamborini, L., Pinto, A., Blondel, A., Minoprio, P., Mozzarelli, A. & De Micheli, C. (2011). *Chemical reviews* **111**, 6919-6946.
- Cortes-Cabrera, A., Gago, F. & Morreale, A. (2015a). *cGRILL software*, <http://farmamol.uah.es>.
- Cortes-Cabrera, A., Gago, F. & Morreale, A. (2015b). *Methods in molecular biology* **1289**, 89-100.
- Coutinho, L., Ferreira, M. A., Cosson, A., Batista, M. M., Batista Dda, G., Minoprio, P., Degrave, W. M., Berneman, A. & Soeiro Mde, N. (2009). *Memorias do Instituto Oswaldo Cruz* **104**, 1055-1062.
- Dunlop, D. S., Neidle, A., McHale, D., Dunlop, D. M. & Lajtha, A. (1986). *Biochem Biophys Res Commun* **141**, 27-32.
- Espaillet, A., Carrasco-Lopez, C., Bernardo-Garcia, N., Pietrosevoli, N., Otero, L. H., Alvarez, L., de Pedro, M. A., Pazos, F., Davis, B. M., Waldor, M. K., Hermoso, J. A. & Cava, F. (2014). *Acta crystallographica. Section D, Biological crystallography* **70**, 79-90.
- Faraci, W. S. & Walsh, C. T. (1988). *Biochemistry* **27**, 3267-3276.
- Faraci, W. S. & Walsh, C. T. (1989). *Biochemistry* **28**, 431-437.
- Fenn, T. D., Holyoak, T., Stamper, G. F. & Ringe, D. (2005). *Biochemistry* **44**, 5317-5327.
- Fey, G., Gould, G. W. & Hitchins, A. D. (1964). *Journal of general microbiology* **35**, 229-236.
- Fisher, G. H., D'Aniello, A., Vetere, A., Padula, L., Cusano, G. P. & Man, E. H. (1991). *Brain Res Bull* **26**, 983-985.
- Fuchs, S. A., Berger, R., Klomp, L. W. & de Koning, T. J. (2005). *Molecular genetics and metabolism* **85**, 168-180.
- Gao, X., Ma, Q. & Zhu, H. (2015). *Applied microbiology and biotechnology*.
- Grishin, N. V., Phillips, M. A. & Goldsmith, E. J. (1995). *Protein Sci* **4**, 1291-1304.
- Hashimoto, A., Nishikawa, T., Oka, T. & Takahashi, K. (1993). *J Neurochem* **60**, 783-786.
- Hellberg, J. E., Matilla, M. A. & Salmond, G. P. (2015). *Frontiers in microbiology* **6**, 137.
- Hemraj S. Nandanwar, G. S. H., Rakesh M. Vohra (2005). *Enzymatic Production of D-Amino Acids*. Humana Press.
- Hoffman, H. E., Jiraskova, J., Cigler, P., Sanda, M., Schraml, J. & Konvalinka, J. (2009). *Journal of medicinal chemistry* **52**, 6032-6041.
- Hubbard, M. A., Thorkildson, P., Welch, W. H. & Kozel, T. R. (2013). *Molecular immunology* **55**, 337-344.
- Kaji, Y., Oshika, T., Takazawa, Y., Fukayama, M. & Fujii, N. (2010). *Chemistry & biodiversity* **7**, 1364-1370.
- Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R. & Losick, R. (2010). *Science* **328**, 627-629.
- Lam, H., Oh, D. C., Cava, F., Takacs, C. N., Clardy, J., de Pedro, M. A. & Waldor, M. K. (2009). *Science* **325**, 1552-1555.
- LLC, S. (2015). *Glide docking program*.
- Major, D. T., Nam, K. & Gao, J. (2006). *Journal of the American Chemical Society* **128**, 8114-8115.
- Shaw, J. P., Petsko, G. A. & Ringe, D. (1997a). *Biochemistry* **36**, 1329-1342.

- Shaw, J. P., Petsko, G. A. & Ringe, D. (1997b). *Biochemistry* **36**, 1329-1342.
- Soda, K., Yoshimura, T. & Esaki, N. (2001). *Chem Rec* **1**, 373-384.
- Spies, M. A. & Toney, M. D. (2003). *Biochemistry* **42**, 5099-5107.
- Spies, M. A. & Toney, M. D. (2007). *Journal of the American Chemical Society* **129**, 10678-10685.
- Spies, M. A., Woodward, J. J., Watnik, M. R. & Toney, M. D. (2004). *Journal of the American Chemical Society* **126**, 7464-7475.
- Stamper, C. G. F., Morollo, A. A. & Ringe, D. (1998). *Biochemistry* **37**, 10438-10445.
- Sun, S. & Toney, M. D. (1999). *Biochemistry* **38**, 4058-4065.
- Watababe, A., Kurokawa, Y., Yoshimura, T., Kurihara, T., Soda, K. & Esaki, N. (1999). *Journal of Biological Chemistry* **274**, 4189-4194.
- Watanabe, A., Yoshimura, T., Mikami, B. & Esaki, N. (1999). *Journal of Biochemistry* **126**, 781-786.
- Watanabe, A., Yoshimura, T., Mikami, B., Hayashi, H., Kagamiyama, H. & Esaki, N. (2002a). *The Journal of biological chemistry* **277**, 19166-19172.
- Watanabe, A., Yoshimura, T., Mikami, B., Hayashi, H., Kagamiyama, H. & Esaki, N. (2002b). *Journal of Biological Chemistry* **277**, 19166-19172.
- Wu, H. M., Kuan, Y. C., Chu, C. H., Hsu, W. H. & Wang, W. C. (2012). *PloS one* **7**, e48301.
- Yang, H., Wang, M., Yu, J. & Wei, H. (2015). *FEMS microbiology letters* **362**.



#### 4.4.1. DISCUSIÓN

La importancia de los D-aminoácidos para la supervivencia de la bacteria así como su implicación en diversas patologías humanas hace que sea necesario el estudio de la estructura y el mecanismo enzimático de las proteínas involucradas en la racemización de los aminoácidos. Se conoce mucho acerca de los inhibidores de las alanina racemasas, algunos son antibióticos naturales y muchos de ellos no son específicos; se ha identificado una nueva familia de inhibidores que no son sustratos de AlaR en *Mycobacterium tuberculosis*. Recientemente, una nueva familia de racemasas localizadas en el periplasma se ha identificado en graves patógenos que afectan a humanos. En contraposición a las enzimas mono específicas (AlaR), las racemasas de amplio espectro (BsR) pueden convertir diez aminoácidos produciendo aminoácidos no canónicos. A través del análisis estructural y bioinformático se han podido determinar las diferencias en la especificidad de sustrato entre las racemasas mono específicas y las de amplio espectro.

Se ha descrito una nueva familia de racemasas de amplio espectro cuyo principal representante es la racemasa de *Vibrio cholerae* BsrV. BsrV puede racemizar aminoácidos alifáticos (no ramificados en  $\beta$ ) y cargados positivamente; sin embargo, no son sustratos los aminoácidos cargados negativamente, aromáticos o alifáticos ramificados en  $\beta$  (Ile, Val y Thr). A través de la resolución estructural de la racemasa de amplio espectro BsrV se han podido identificar las características estructurales que definen esta nueva familia de racemasas. A pesar de que el plegamiento de BsrV es similar al de las AlaR, se observan diferencias importantes al superponer ambas estructuras, así la disposición de los dominios que presenta BsrV es diferente con respecto a las AlaR, estas diferencias se deben a la presencia de una hélice  $\alpha$  extra y una extensión en el N-terminal. La inserción en el N-terminal, así como la presencia de un lazo corto que conforma la entrada al sitio catalítico, hace que el canal de entrada sea más ancho en la BsrV y que la separación de los sitios activos sea mayor en BsrV comparada con AlaR.

Las estructuras a alta resolución ya resueltas junto con los análisis bioinformáticos aportan información clave sobre cómo es la unión del sustrato y la catálisis enzimática de estas racemasas. El sitio activo de BsrV presenta diferencias determinantes que permiten su plasticidad a la hora de racemizar aminoácidos. La presencia de una Pro en sustitución de la Tyr que presentan las AlaR en la entrada del sitio catalítico, hace que la cavidad se ensanche permitiendo la entrada de sustratos de cadena lateral más grande. Además, la BsrV carece de la lisina carbamilada de manera que la Arg que participa en la unión a sustrato está estabilizada mediante un ión  $\text{Cl}^-$ . A través de los mapas generados mediante el programa cGRILL (Cortes-Cabrera *et al.*, 2015) se ha confirmado que la cavidad catalítica de la BsrV es mucho más amplia que la que presentan las AlaR. Mediante los programas cGRILL y Glide (LLC, S. (2015). *Glide docking program.*) se comprobó que, efectivamente, los aminoácidos cargados negativamente (Glu y Asp) no pueden ser sustratos de BsrV debido a la repulsión de carga existente en el sitio activo. Para

determinar las diferencias estructurales que definen la especificidad por el sustrato en las racemasas mono y multi específicas, se realizó un estudio de *docking* usando la estructura de la racemasa de *Bacillus stearothermophilus* (AlaR<sub>Bs</sub>), en complejo con la D-Ala, como modelo. La estructura de BsrV se superpuso sobre este modelo y se reemplazó el sustrato (D-Ala) por todos los tipos de aminoácidos posibles, los mismos estudios se realizaron de manera paralela con AlaR. Se vio que en los aminoácidos alifáticos (Ala, Ser, Cys, Leu, Gln y Met) las cadenas laterales se estabilizan mediante interacciones hidrofóbicas con la Met347, los aminoácidos básicos (Arg, Lys y Orn) están además estabilizados por un puente de hidrógeno con la Tyr394 y mediante interacciones electrostáticas con el Asp268 y el fosfato del PLP, que además está de acuerdo con los mapas de afinidad calculados mediante cGRILL. Además, los estudios de *docking* permiten determinar cómo las BsR, y no las AlaR, pueden racemizar una gran variedad de aminoácidos. La sustitución de la Tyr394 en AlaR<sub>Bs</sub> por la Pro391 en BsrV (un residuo de Ala en otras BsR) proporciona un ensanchamiento del sitio activo y permite la entrada de varios sustratos con cadenas laterales diferentes, mientras que en AlaR los mismos sustratos generan múltiples choques con los aminoácidos que conforman la cavidad catalítica.

Considerando el papel que juegan los NCDAA en la regulación de la pared celular y el metabolismo bacteriano, estos conocimientos establecen las bases para el diseño de drogas nuevas y más efectivas.

Las conclusiones de este trabajo se resumen a continuación:

- ✓ Se han definido los diferentes tipos de enzimas PLP-dependientes en base a su plegamiento estructural.
- ✓ Se ha descrito el mecanismo catalítico de AlaR y comparado con el de la BsrV.
- ✓ Se han establecido las bases estructurales que definen a las racemasas monoespecíficas y a las racemasas de amplio espectro .
- ✓ Se ha descrito una nueva familia de racemasas de amplio espectro dentro de la cual se encuentra la racemasa de *Vibrio cholerae* BsrV.
- ✓ Mediante la caracterización estructural y los estudios bioinformáticos se han podido explicar las diferencias en la selectividad de sustrato entre BsrR y AlaR.

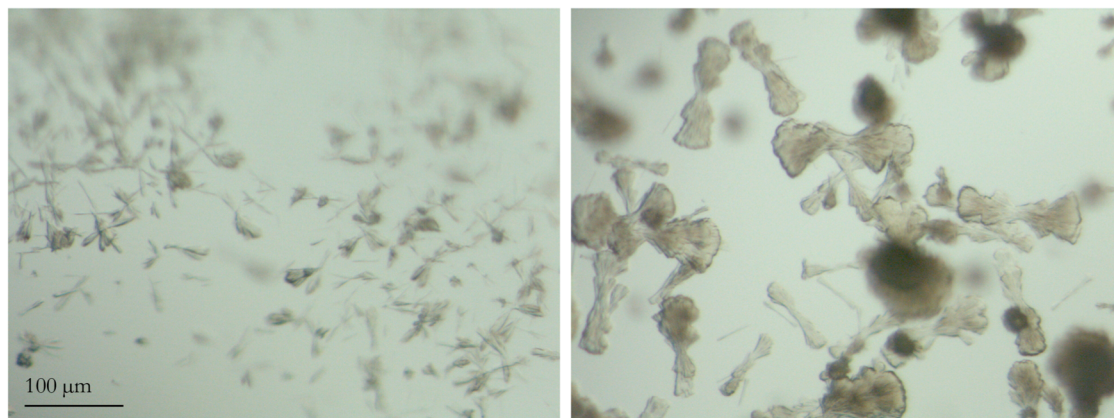
## 5. MATERIAL INÉDITO



## 5.1. ANEXO I. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA MONOESPECÍFICA DE *Bacillus subtilis* Bsu17640

### I.1. CRISTALIZACIÓN

Se realizó un muestreo masivo de condiciones de cristalización, se obtuvieron cristales en varias condiciones (Tabla I.1), todos ellos presentaban macla (Fig. I.1).



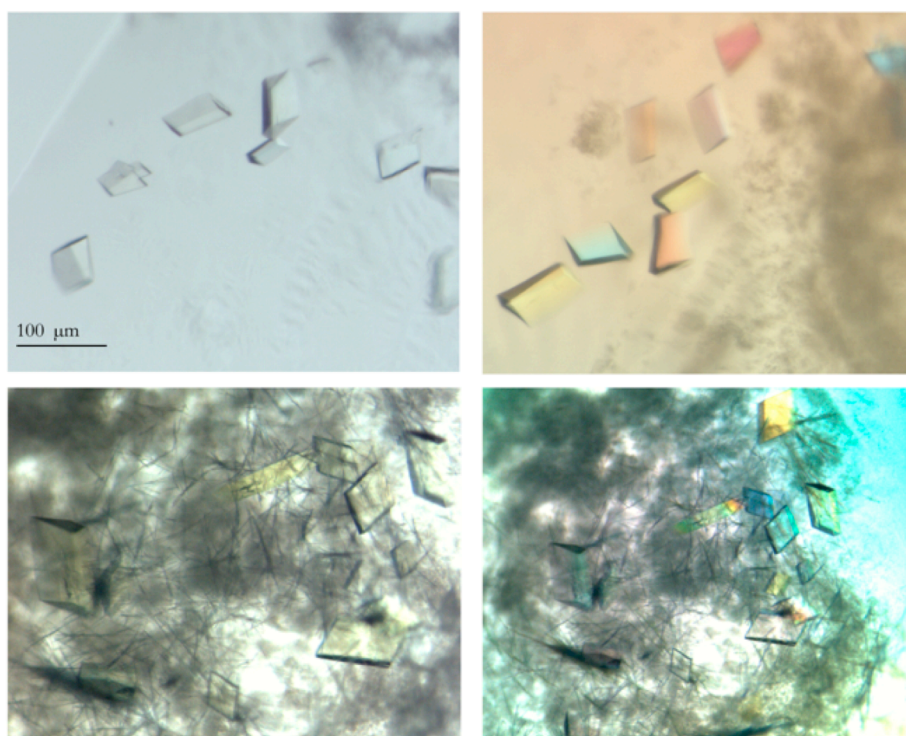
**Figura I.1. Cristales iniciales de Bsu17640.** A la derecha se muestra una ampliación de la gota de cristalización.

**Tabla I.1.** Condiciones de cristalización iniciales de Bsu17640

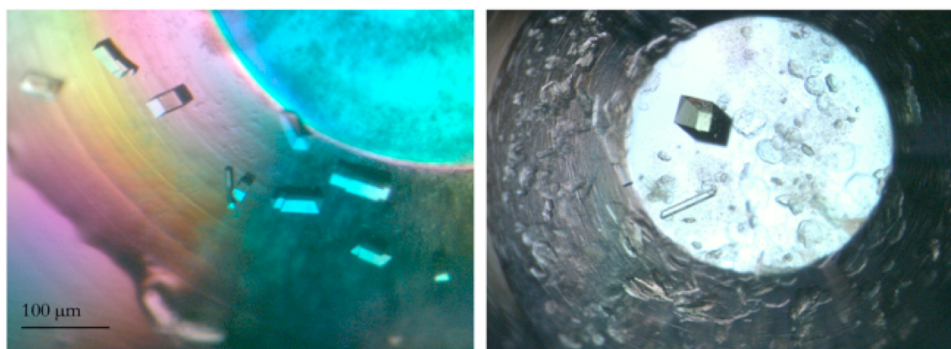
Proteína	Precipitante	Tampón*	Aditivo*
<b>Bsu17640</b>	20% (v/v) PEG 400	HEPES pH 7,5	CaCl <sub>2</sub>
7mg/ml en 20 mM Tris-HCl,	30% (m/v) PEG 8000	Cacodilato de sodio pH 6,5	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
200 mM NaCl	8% (m/v) PEG 8000	Tris-HCl pH 8,5	—
pH 7,5	10% (v/v) PEG MME 5000	HEPES pH 7,0	5% Tacsinato

\*Todos los tampones y aditivos se encuentran a 100 mM y 200 mM, respectivamente.





**Figura I.2. Cristales de Bsu17640 con diferentes detergentes.** Se muestran imágenes de cristales de proteína sin polarizador (izquierda) y con polarizador (derecha).



**Figura I.3. Cristales optimizados de Bsu17640 obtenidos bajo aceite.** En la fotografía de la derecha se muestra una ampliación de un cristal obtenido mediante la técnica de *microbatch*.

Se llevó a cabo la optimización de los cristales obtenidos variando los porcentajes de PEG 4000 y el valor de pH así como la concentración de la proteína y el volumen de las gotas, estos ensayos se realizaron usando la técnica de difusión de vapor así como la técnica de *microbatch*. También se analizó el efecto de los detergentes en la cristalización (*Detergent screen*, Hampton Research, Aliso Viejo, California, USA) (Fig. I.2). Los cristales optimizados crecen hasta alcanzar su tamaño óptimo en menos de una semana (Fig. I.3). Aunque todos los cristales obtenidos después de la optimización eran grandes y con

buen hábito cristalino los que dieron mejor resultado fueron los obtenidos bajo la técnica de *microbatch* y usando una solución crioprotectora de 70% (v/v) Paratona.

## I.II. DETERMINACIÓN ESTRUCTURAL

Los cristales de Bsu17640 se midieron en el sincrotrón de Grenoble (ESRF) obteniendo un conjunto de datos de difracción a 2,1 Å de resolución. Los cristales pertenecen al sistema cristalino tetragonal, P<sub>4</sub><sub>3</sub>2, con una dimensiones de celdilla de  $a = 141,09$  Å,  $b = 290,41$  Å,  $c = 87,21$  Å,  $\alpha = \beta = \gamma = 90^\circ$ . En la Tabla I.2 se recogen los resultados obtenidos del procesado de imágenes.

**Tabla I.2.** Estadísticas del procesado de datos de difracción.

	<b>Bsu 17640</b>
<b>Sistema cristalino</b>	Tetragonal
<b>Grupo espacial</b>	P <sub>4</sub> <sub>3</sub> 2
<b><math>a</math> (Å)</b>	14,09
<b><math>b</math> (Å)</b>	290,41
<b><math>c</math> (Å)</b>	87,21
<b><math>\alpha</math> (°)</b>	90
<b><math>\beta</math> (°)</b>	90
<b><math>\gamma</math> (°)</b>	90
<b>Fuente de radiación</b>	Sincrotrón
<b>Longitud de onda (Å)</b>	0,980110
<b>Intervalo de resolución (Å)</b>	49,15–2,1 (2,175–2,1)
<b>Observaciones</b>	1074403
<b>Reflexiones únicas</b>	5253
<b>Compleitud</b>	100 (100)
<b>Multiplicidad</b>	11,4
<b>*<math>R_{merge}</math></b>	0,096 (0,55)
<b><math>\langle I/\sigma(I) \rangle</math></b>	21,5 (4,8)

Los valores entre paréntesis corresponden a la resolución más alta.

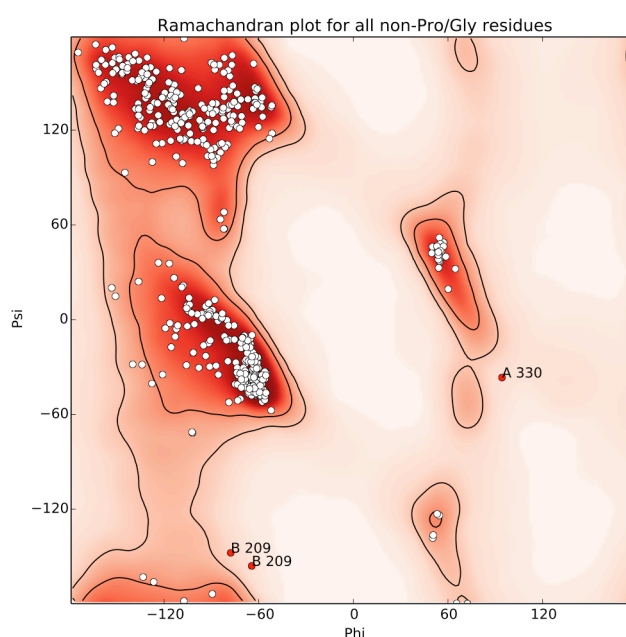
\* $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$ , donde  $I_i(hkl)$ .

La estructura se resolvió mediante el método de Remplazo Molecular utilizando el programa MOLREP y usando la alanina racemasa de *Bacillus anthracis* (código PDB 3HA1) (Counago *et al.*, 2009) como modelo inicial. Esta proteína presenta un 64% de homología y un 43% de identidad, respectivamente. Mediante el cálculo del coeficiente de Matthews se pudo estimar el porcentaje de solvente del cristal (51,61%), así como la presencia de dos moléculas en la unidad asimétrica ( $V_M = 2,57$  Å<sup>3</sup>Da<sup>-1</sup>) (Matthews, 1968).

Se obtuvo un modelo final de Bsu17640 refinado a 2,1 Å de resolución. El mapa de densidad electrónica obtenido permitió modelar la cadena polipeptídica del residuo 1 al 386, 2 moléculas de PLP, 3 iones cloro, 5 moléculas de TRIS, 5 moléculas de PEG y 505 moléculas de agua. Las estadísticas del refinamiento quedan recogidas en la Tabla I.3. El modelo final se validó comprobando que todos los residuos presentan conformaciones permitidas como se muestra en el diagrama de Ramachandran (Figura I.4).

**Tabla I.3.** Estadísticas de refinamiento para el modelo de Bsu17640.

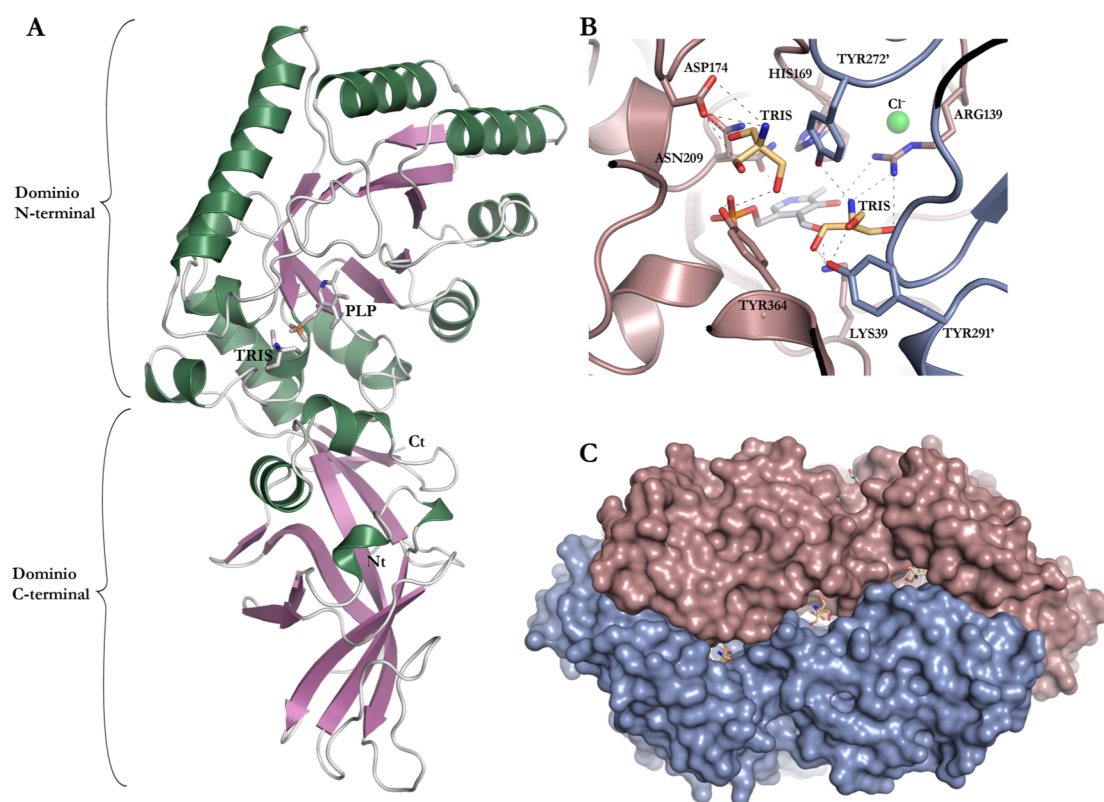
Parámetros	Bsu 17640
$R_{\text{work}}/R_{\text{free}}$ (%)	17,45/ 21,70
RMSD de los enlaces (Å)	2,48
RMSD de los ángulos (°)	0,034
Factor <i>B</i> medio (Å <sup>2</sup> )	23,8
Número de átomos	
Proteína	6039
Ligando	118
Agua	504
Mapa de Ramachandran	
Favorecidos (%)	97
Permitidos (%)	2,5
Prohibidos (%)	0,52



**Figura I.4. Validación de la estructura Bsu17640.** Diagrama de Ramachandran generado con PHENIX.

### I.III. ANÁLISIS ESTRUCTURAL

La estructura de Bsu17640 es globular y está constituida por dos dominios, en el extremo amino presenta un barril  $\alpha/\beta$  de ocho hebras  $\beta$ , en el cual se encuentra el sitio de unión a fosfato; y en el extremo carboxilo encontramos un dominio de láminas  $\beta$  (Fig. I.5 a). El sitio activo de Bsu17640 es muy similar al de la racemasas monoespecífica AlaR previamente descrita. En el caso de Bsu17640 el cofactor PLP no se encuentra covalentemente unido a la lisina (Lys39). La proteína Bsu17640 cristalizó en una condición que presentaba TRIS pH 8,5 como tampón; la calidad del mapa de densidad permitió identificar varias moléculas de TRIS en la estructura, dos de ellas se encuentran estabilizadas por diversas interacciones de hidrógeno y polares en el sitio activo de la proteína (Fig. I.5 b); también se ha encontrado una molécula de TRIS con doble conformación en el centro del canal que conecta ambos sitios activos (Fig. I.5 c).



**Figura I.5. Estructura tridimensional de la racemasa Bsu17640.** a) Representación del plegamiento de un monómero de Bsu17640 con el dominio N-terminal formado por un barril  $\alpha/\beta$  y el dominio C-terminal constituido por láminas  $\beta$ . b) Detalle del sitio activo. Se muestran los residuos que lo forman y las interacciones presentes con el cofactor (PLP) y el ligando TRIS. c) Superficie molecular del dímero de Bsu17640. Se muestra cada uno de los monómeros en color azul y marrón, así como las moléculas de TRIS que están presentes en los sitios activos y en el canal que conecta ambos.

El trabajo experimental relativo a esta proteína está completado, se está realizando un análisis bioinformático que permitirá conocer en profundidad los detalles del sitio activo de Bsu17640 y

comprender el mecanismo catalítico de esta enzima que no presenta el cofactor PLP unido covalentemente a la Lys. Actualmente se está trabajando en la puesta en común de los resultados para la elaboración del artículo de investigación que se pretende publicar en la correspondiente revista de carácter científico.



## 5.2. ANEXO II. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA DE AMPLIO ESPECTRO DE *Acinetobacter baumannii* BsrAb

### II.I. CRISTALIZACIÓN

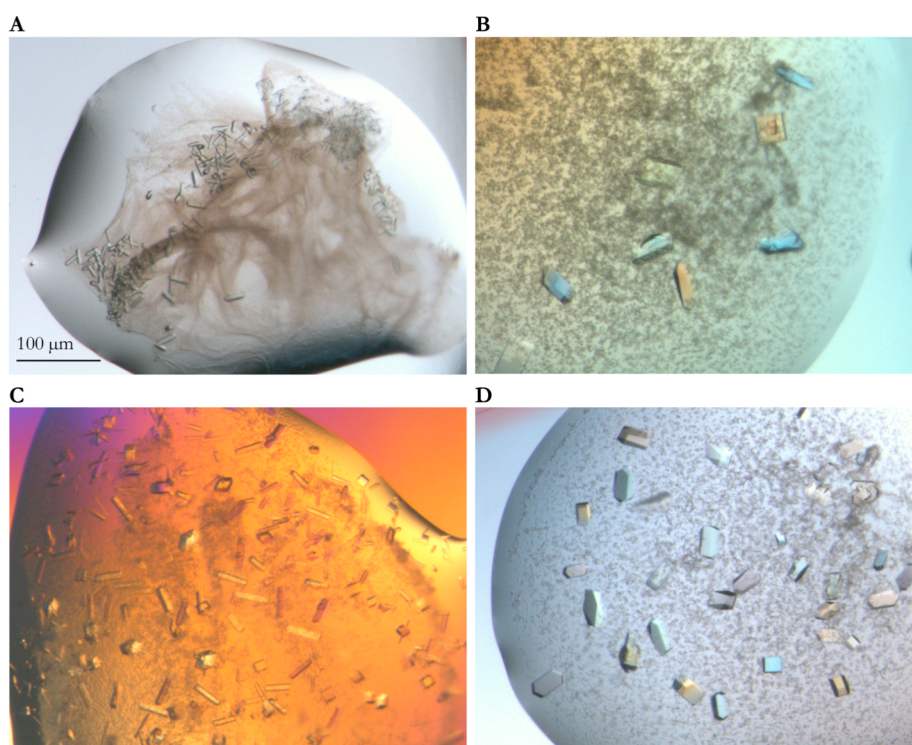
Tras el muestreo masivo de condiciones de cristalización se obtuvieron cristales en varias condiciones (Tabla II.1), con buen tamaño y hábito cristalino (Fig. II.1).

**Tabla II.1. Condiciones de cristalización iniciales de BsrAb**

Proteína	Precipitante	Tampón*	Aditivo*
<b>BsrAb</b>	25% (m/v) PEG 1500	MIB** tampón pH 5,0	—
13 mg/ml en 20 mM Tris-HCl pH 7.5	20% (m/v) PEG 6000	Acetato de sodio pH 5,0	CaCl <sub>2</sub>
	25% (m/v) PEG 1500	SPG** tampón pH 4 y 5	—
	10% (m/v) PEG 3000	Acetato de sodio pH 4,5	Zn(O <sub>2</sub> CCH <sub>3</sub> ) <sub>2</sub>

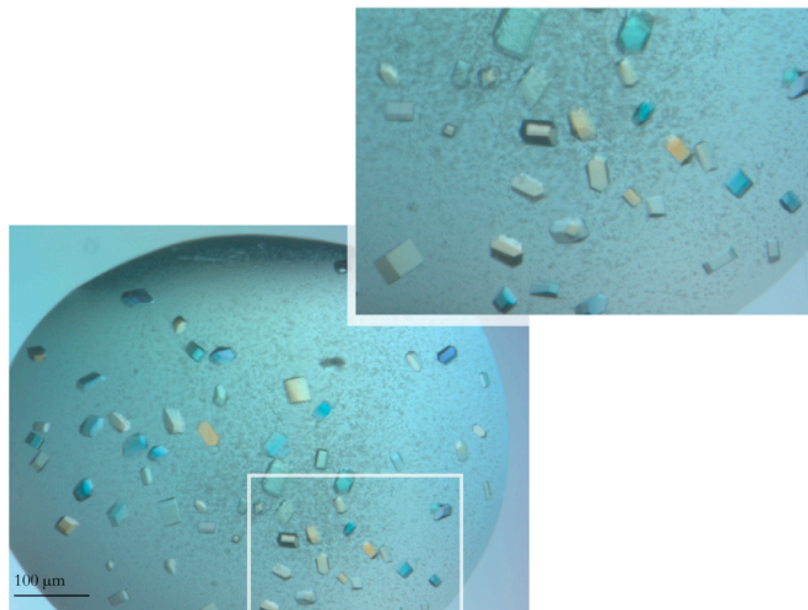
\*Todos los tampones y aditivos se encuentran a 100 mM y 200 mM, respectivamente.

\*\* Tampones MIB y SPG están compuestos de malonato, imidazol y ácido bórico a un ratio molar de 2:3:3 y ácido succínico, fosfato sódico dihidrógeno y glicina a un ratio molar 2:7:7, respectivamente.



**Figura II.1. Condiciones comerciales donde cristalizó la proteína BsrAB.** **a)** 10% (m/v) PEG 3000, 0,2 M acetato de zinc y 0,1 acetato de sodio pH 4,5. **b)** 25% (m/v) PEG 1500 y 0,1 M tampón SPG pH 5. **c)** 25% (m/v) PEG 1500 y 0,1 M tampón SPG pH 4. **d)** 25% (m/v) PEG 1500 y 0,1 M tampón MIB pH 5.

La optimización de los cristales obtenidos se llevó a cabo variando los porcentajes de PEG 4000 y el valor de pH así como la concentración de la proteína y el volumen de las gotas, estos ensayos se realizaron usando la técnica de difusión de vapor así como la técnica de *microbatch*. Los cristales optimizados se forman en tres días en 25% (m/v) PEG 1500 tampón MIB pH 4, 5 y 6 en difusión de vapor (Fig. II.2). Aunque los cristales tenían buen aspecto hubo que probar diferentes crioprotectores para mejorar la difracción de los mismos, dando como mejor resultado el uso de 70% Paratona.



**Figura II.2. Cristales optimizados de BsrAb.** Fotografía de una gota de cristalización usando un polarizador, arriba a la derecha se muestra una ampliación de la misma.

## II.II. DETERMINACIÓN ESTRUCTURAL

Los experimentos de difracción de rayos-X se llevaron a cabo en el sincrotrón ALBA (Barcelona) obteniendo un conjunto de datos de difracción a 1,8 Å de resolución. Los cristales pertenecen al sistema cristalino ortorrómbico, C222<sub>1</sub> con una dimensiones de celdilla de  $a = b = 72,88 \text{ Å}$ ,  $c = 332,83 \text{ Å}$ ,  $\alpha = \beta = \gamma = 90^\circ$ . En la Tabla II.2 se recogen los resultados obtenidos del procesado de imágenes.

Tabla II.2. Estadísticas del procesado de datos de difracción.

	<b>BsrAb</b>
<b>Sistema cristalino</b>	Ortorrómbico
<b>Grupo espacial</b>	C222 <sub>1</sub>
<i>a</i> (Å)	142,0
<i>b</i> (Å)	290,5
<i>c</i> (Å)	87,3
$\alpha$ (°)	90
$\beta$ (°)	90
$\gamma$ (°)	90
<b>Fuente de radiación</b>	Sincrotrón
<b>Longitud de onda (Å)</b>	1.000020
<b>Intervalo de resolución (Å)</b>	41,2–2,4 (2,43–2,4)
<b>Observaciones</b>	464967
<b>Reflexiones únicas</b>	72269
<b>Complejidad</b>	95,8 (100)
<b>Multiplicidad</b>	6,4
<b>*R<sub>merge</sub></b>	0,123 (0,54)
<b><math>\langle I/\sigma(I) \rangle</math></b>	12,5 (5,6)

Los valores entre paréntesis corresponden a la resolución más alta.

\*R<sub>merge</sub> =  $\sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$ , donde  $I_i(hkl)$ .

La estructura tridimensional se resolvió mediante el método de Remplazo Molecular usando el programa MOLREP con la racemasa de amplio espectro de *Vibrio cholerae* (código PDB 4BEU) (Espaillat *et al.*, 2014) como modelo estructural, con la que presentaba un 70% y un 47% de homología y de identidad, respectivamente. Mediante el cálculo del coeficiente de Matthews se pudo estimar el porcentaje de solvente del cristal (50,43%), así como la presencia de cuatro moléculas en la unidad asimétrica ( $V_M = 2,48 \text{ Å}^3 \text{Da}^{-1}$ ) (Matthews, 1968).

El modelo final de BsrAb refinado se obtuvo a 2,4 Å de resolución. El mapa de densidad electrónica obtenido permitió modelar la cadena polipeptídica del residuo 19 al 390, 4 moléculas de PLP, 5 iones cloro, 7 moléculas de PEG y 623 moléculas de agua. Las estadísticas del refinamiento quedan recogidas en la Tabla II.3. El modelo final se validó comprobando que todos los residuos presentan conformaciones permitidas como se muestra en el diagrama de Ramachandran (Fig. II.3).



Tabla II.3. Estadísticas de refinamiento para el modelo de BsrAb.

Parámetros	BsrAb
$R_{\text{work}}/R_{\text{free}}$ (%)	18,39/25,00
RMSD de los enlaces (Å)	0,016
RMSD de los ángulos (°)	1,50
Factor <i>B</i> medio (Å <sup>2</sup> )	31,62
Número de átomos	
Proteína	1444
Ligando	150
Agua	623
Mapa de Ramachandran	
Favorecidos (%)	96
Permitidos (%)	2,8
Prohibidos (%)	0,84

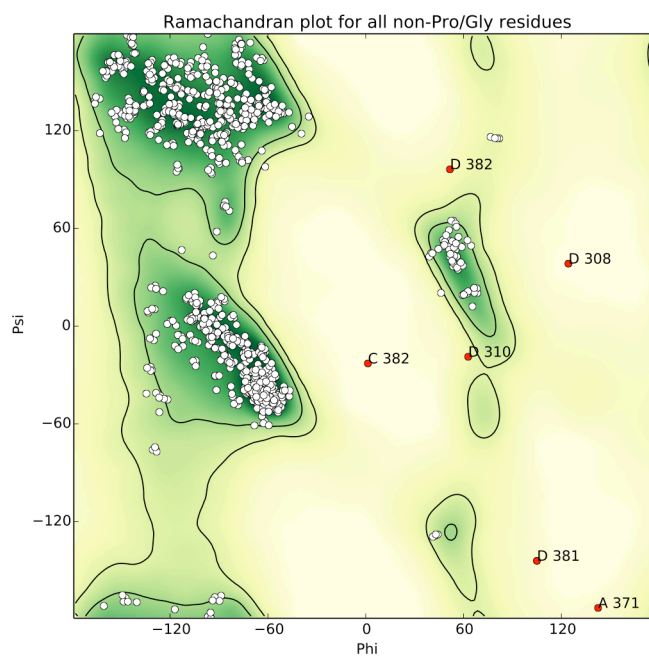
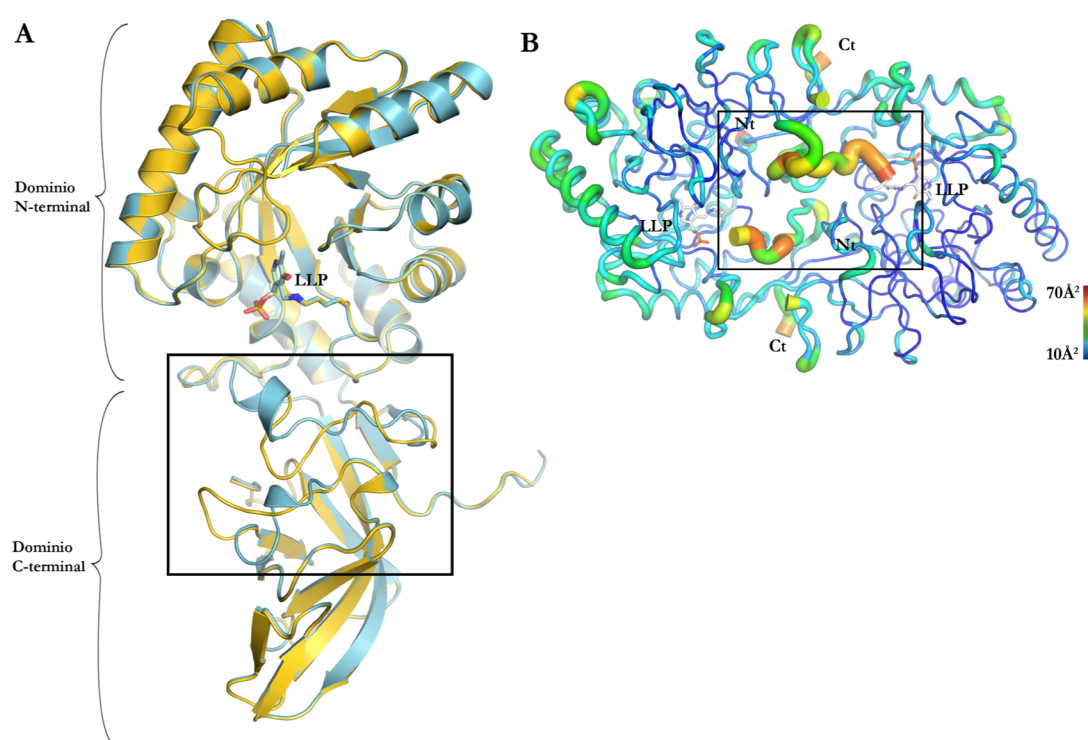


Figura II.3. Validación de la estructura BsrAb. Diagrama de Ramachandran generado con el programa PHENIX.

### II.III. ANÁLISIS ESTRUCTURAL

La proteína BsrAb presenta dos dominios diferenciados N-terminal y C-terminal. El dominio N-terminal que contiene el sitio de unión a PLP se compone de un barril  $\alpha/\beta$  de ocho hebras  $\beta$  y el dominio C-terminal presenta láminas  $\beta$  (Fig. II.4 a). Los monómeros de las racemasas de amplio espectro resueltas hasta el momento presenta el mismo plegamiento; sin embargo, en BsrAb los dominios C-terminal de los monómeros que componen el dímero funcional presentan diferencias en el plegamiento de dos bucles que se encuentran cerca del sitio activo (Fig. II.4 a). Además, se analizaron los factores de agitación térmica y se observó que los valores para las zonas que conectan ambos sitios activos son muy altos, confirmandose la movilidad que presentan dichos bucles (Fig. II.4 b).



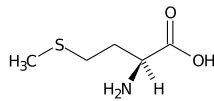
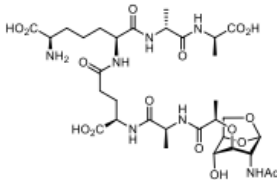
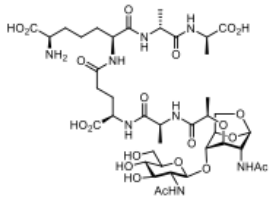
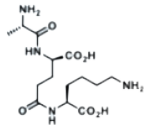
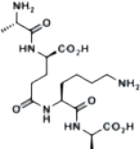
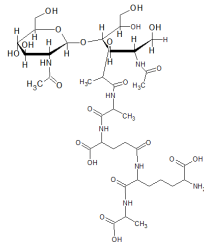
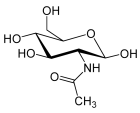
**Figura II.4. Estructura tridimensional de BsrAb.** a) Superposición de los monómeros que componen BsrAb, se identifica el dominio N-terminal donde se encuentra el sitio de unión a PLP y el dominio C-terminal, el área recuadrada señala la zona en la que los bucles de ambos monómeros presentan conformaciones diferentes. b) Representación de los factores de agitación térmica para BsrAb. Los valores cercanos al azul reflejan baja agitación térmica, mientras que si se acercan al rojo la agitación térmica es mayor, lo que implica más flexibilidad. La zona que coincide con los bucles móviles que conectan ambos sitios activos está resaltada dentro del cuadro negro.

### II.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS

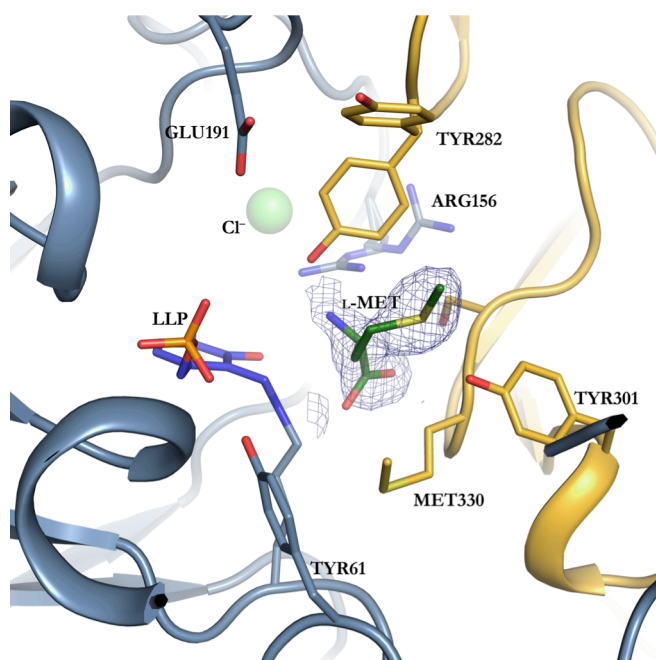
Para intentar entender el mecanismo de reacción de la enzima se realizaron una serie de experimentos de cristalización en complejo con diferentes sustratos de la enzima. Las técnicas que se emplearon fueron la co-cristalización de la proteína con el ligando y la inmersión (ver apartado de

Técnicas Cristalográficas). Los ligandos que se usaron para la cristalización junto con la proteína fueron muy diversos, desde NCDAA (L-Met) hasta análogos de PG tanto naturales como sintéticos. En la Tabla II.4 se listan algunos de los ligandos empleados.

**Tabla II.4. Ligandos empleados para la cristalización con BsrAb**

Ligando	Fórmula química	Estructura	Tipo
<b>L-Met</b>	$C_5H_{11}NO_2S$		Aminoácido
<b>DHL-51</b>	$C_{32}H_{51}N_7O_{16}$		Análogo sintético de PG G(-)
<b>DHL-89</b>	$C_{40}H_{64}N_8O_{21}$		Análogo sintético de PG G(-)
<b>DHC-33</b>	$C_{15}H_{22}N_4O_6$		Análogo sintético de PG G(+)
<b>DHC-34</b>	$C_{17}H_{27}N_5O_7$		Análogo sintético de PG G(+)
<b>M4</b>	$C_{37}H_{63}N_7O_{21}$		Producido a partir de la degradación del PG de <i>E.coli</i>
<b>NAG</b>	$C_8H_{15}NO$		Unidad esencial de PG

Se obtuvieron cristales mediante inmersión con L-Met que se midieron en el sincrotrón ALBA (Barcelona). En la figura II.5 se muestra un detalle de la calidad del mapa de densidad  $2F_o - F_c$  para este ligando. Además, se observa que la Tyr282 presenta una doble conformación lo que podría indicar que estaría actuando como tapadera en la entrada al sitio activo de la enzima.



**Figura II.5. Mapa de densidad electrónica para el complejo BsrAb con L-Met.** Cada uno de los monómeros que forman el sitio activo aparece coloreado en amarillo y azul, se muestran como varillas los residuos que interaccionan con el sustrato, la L-Met y el LLP; el cloruro aparece como una esfera verde. La densidad electrónica corresponde al mapa  $2F_o - F_c$  a  $1\sigma$ .

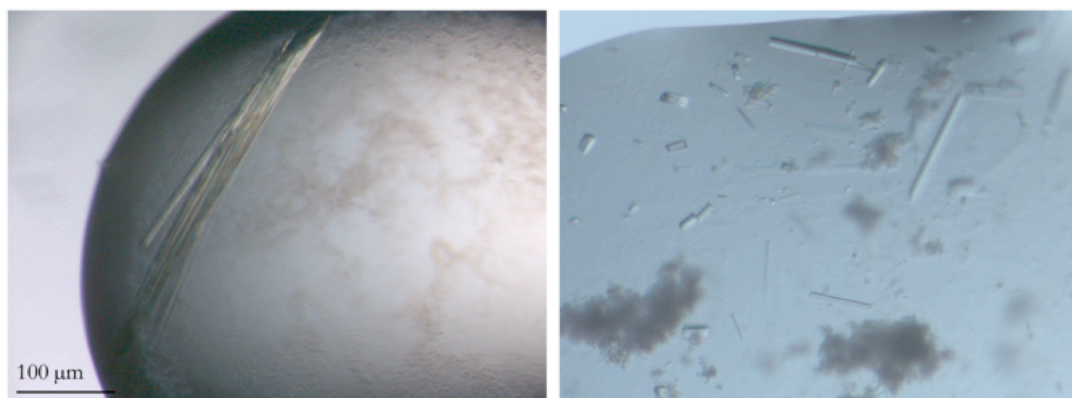
Además de los ensayos de co-cristalización ya mencionados, se pretende cristalizar la proteína en presencia de otros aminoácidos (Arg, Leu, Asn, His, Gln y Lys). Se han obtenido datos de difracción de BsrAb en presencia de análogos de PG que están siendo analizados actualmente. Estos resultados junto con los análisis bioinformáticos pertinentes permitirán la elaboración de un trabajo de investigación apto para publicarse en una revista científica.



### 5.3. ANEXO III. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA RACEMASA DE AMPLIO ESPECTRO DE *Kingella oralis* BsrKo

#### III.I. CRISTALIZACIÓN

Se obtuvieron varias condiciones de cristalización en los muestreos iniciales que presentaban cristales (Tabla III.1), todos ellos estaban maclados (Fig. III.1).

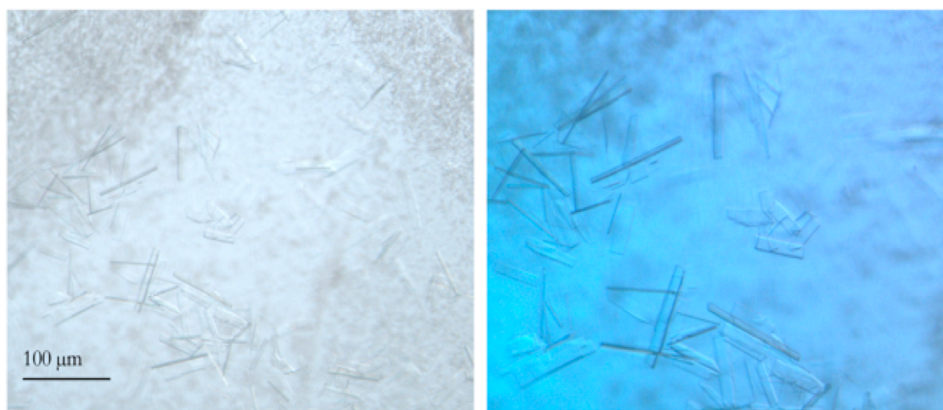


**Figura III.1. Cristales de BsrKo.** a) Gota de cristalización inicial con cristales maclados. Condición de cristalización: 30% (v/v) PEG MME 2000 y 0,1 M Tiocianato de potasio. b) Cristales únicos en forma de prisma rectangular de pequeño tamaño obtenidos a partir de variaciones en la condición de cristalización inicial (30% (v/v) PEG MME 2000 y 0,1 M Tiocianato de potasio).

**Tabla III.1. Condiciones de cristalización iniciales de BsrKo**

Proteína	Precipitante	Tampón*	Aditivo*
<b>BsrKo</b> 7 mg/ml en 20 mM Tris-HCl pH 7,5	30% (v/v) PEG MME 2000	—	100 mM K/Tiocianato
	25% (m/v) PEG 3350	Bis-Tris pH 5,5	NaCl
	25% (m/v) PEG 3350	Bis-Tris pH 5,5	MgCl <sub>2</sub>
	30% (m/v) PEG 4000	Acetato de sodio pH 4,6	Acetato amónico

\*Todos los tampones y aditivos se encuentran a 100 mM y 200 mM, respectivamente.



**Figura III.2. Cristales optimizados de BsrKo.** Fotografía de cristales con polarizador (derecha) y sin él (izquierda).

La optimización de los cristales obtenidos se llevó a cabo variando los porcentajes de PEG y la concentración de la proteína en difusión de vapor. Además, se realizó un muestreo de detergentes. Los mejores cristales se forman a las dos semanas en 34% (v/v) PEG MME 2000, 100 mM K/ Tiocianato, las gotas se prepararon mezclando 2  $\mu\text{L}$  de reservorio con 2  $\mu\text{L}$  de proteína y 0,5  $\mu\text{L}$  de aditivo CYMAL<sup>®</sup>-5; se empleó la técnica de *microbatch* (Fig. III.2). El crioprotector utilizado fue 20% PEG 400.

### III.II. DETERMINACIÓN ESTRUCTURAL

La recogida de los datos de difracción se llevó a cabo en el sincrotrón ALBA (Barcelona) obteniendo un conjunto de datos a 1,94 Å de resolución. Los cristales pertenecen al sistema cristalino ortorrómbico,  $P2_12_12_1$  con una dimensiones de celdilla de  $a = 48,51$  Å,  $b = 126,48$  Å,  $c = 150,49$  Å,  $\alpha = \beta = \gamma = 90^\circ$ . En la Tabla III.2 se recogen los resultados obtenidos del procesado de imágenes.

Tabla III.2. Estadísticas del procesado de datos de difracción.

Parámetros	BsrKo
Sistema cristalino	Ortorrómbico
Grupo espacial	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
<i>a</i> (Å)	48,51
<i>b</i> (Å)	126,48
<i>c</i> (Å)	150,49
$\alpha$ (°)	90
$\beta$ (°)	90
$\gamma$ (°)	90
Fuente de radiación	Sincrotrón
Longitud de onda (Å)	0,97948
Intervalo de resolución (Å)	75,24 – 1,94 (1,99 – 1,94)
Observaciones	785550
Reflexiones únicas	69161
Complejidad	100 (99,8)
Multiplicidad	11.4
* <i>R<sub>merge</sub></i>	0,17 (0,42)
$\langle I/\sigma(I) \rangle$	14,8 (5,1)

Los valores entre paréntesis corresponden a la resolución más alta.

\* $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$ , donde  $I_i(hkl)$ .

La estructura tridimensional se resolvió mediante el método de Remplazo Molecular usando el programa MOLREP con la racemasa de amplio espectro de *Proteus mirabilis* (código PDB 4FS9) (Wu *et al.*, 2012) como modelo estructural, con la que presentaba un 72% de homología y un 55% de identidad, respectivamente. Mediante el cálculo del coeficiente de Matthews se pudo estimar el porcentaje de solvente del cristal (52,58%), así como la presencia de dos moléculas en la unidad asimétrica ( $V_M = 2,59 \text{ Å}^3 \text{ Da}^{-1}$ ) (Matthews, 1968).

El modelo final de BsrKo refinado se obtuvo a 1,94 Å de resolución. El mapa de densidad electrónica obtenido permitió modelar la cadena polipeptídica del residuo 25 al 403, 7 iones potasio, 2 moléculas de PLP, 2 moléculas de isotiocianato, 22 moléculas de PEG y 567 moléculas de agua. Las estadísticas del refinamiento quedan recogidas en la Tabla III.3. El modelo final se validó comprobando que todos los residuos presentan conformaciones permitidas como se muestra en el diagrama de Ramachandran (Fig. III.3).



Tabla III.3. Estadísticas de refinamiento para el modelo de BsrKo.

Parámetros	BsrKo
$R_{\text{work}}/R_{\text{free}}$ (%)	15,32/ 17,97
RMSD de los enlaces (Å)	0,014
RMSD de los ángulos (°)	1,38
Factor <i>B</i> medio (Å <sup>2</sup> )	17,62
Número de átomos	
Proteína	5917
Ligando	164
Agua	567
Mapa de Ramachandran	
Favorecidos (%)	97,5
Permitidos (%)	2,2
Prohibidos (%)	0,3

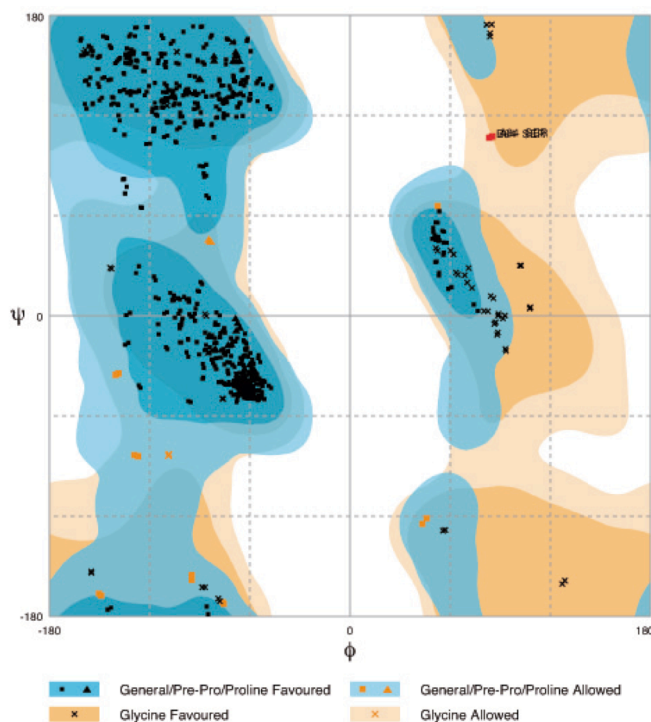


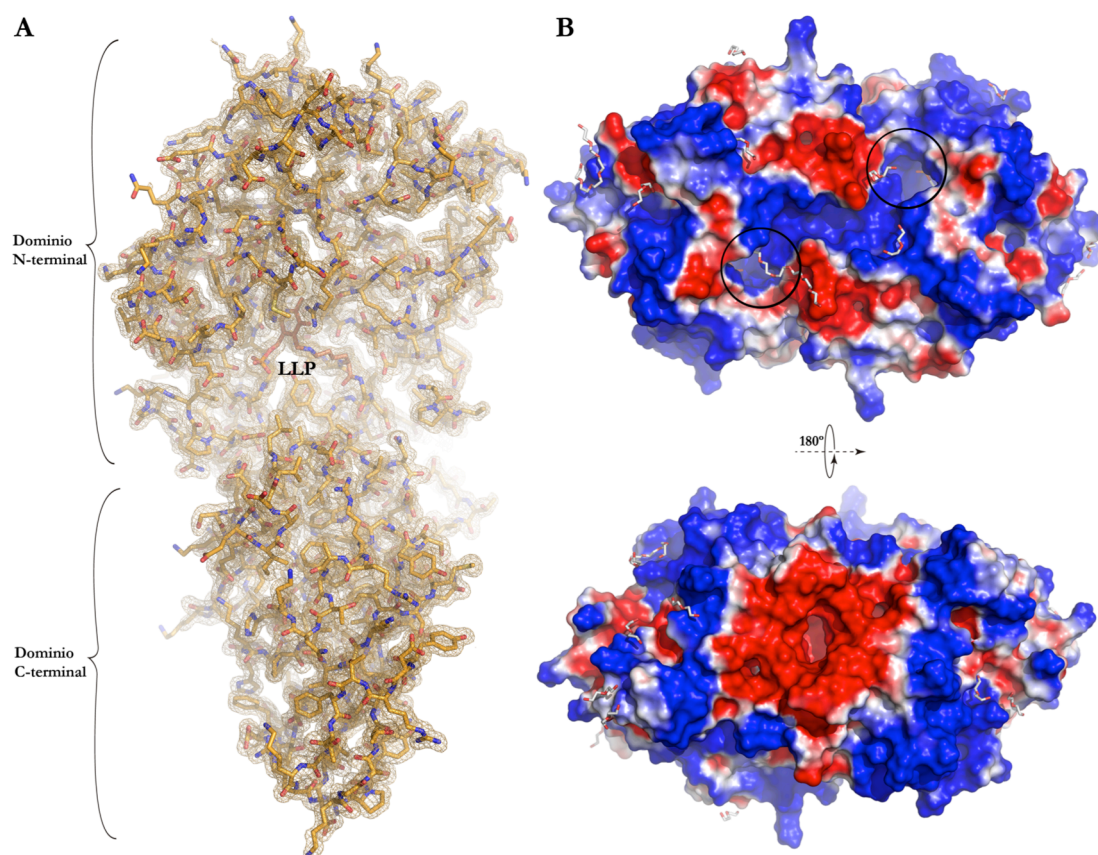
Figura III.3. Validación de la estructura BsrKo. Diagrama de Ramachandran generado con PROCHECK.

### III. III. ANÁLISIS ESTRUCTURAL

La proteína BsrKo presenta el plegamiento típico de las racemasas PLP dependientes, con dos dominios N-terminal y C-terminal diferenciados. En el caso de esta racemasa de amplio espectro, el dímero formado se compone de dos monómeros idénticos en cuanto al plegamiento de los mismos se

refiere. En la figura III.4a se indican los dominios en los monómeros que conforman la estructura homodimérica de BsrKo, asimismo se puede apreciar la calidad del mapa de densidad  $2F_o - F_c$ .

A través de la aplicación APBS que se encuentra contenida dentro del programa PyMol se pudo calcular el potencial electrostático de la proteína; se observa que, tanto las entradas a los sitios activos como el canal que conecta ambos es altamente básico, mientras que la cara opuesta presenta un marcado carácter ácido (Fig. III.4 b).



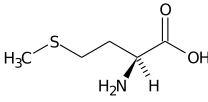
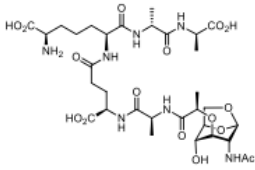
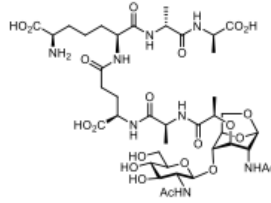
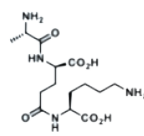
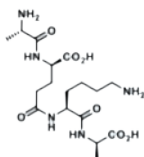
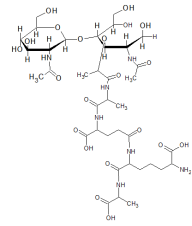
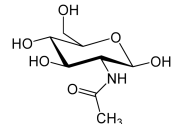
**Figura III.4. Plegamiento general y potencial electrostático de BsrKo.** a) Mapa de densidad para uno de los monómeros que compone el homodímero BsrKo. La densidad electrónica corresponde al mapa  $2F_o - F_c$  a  $1\sigma$ . b) Representación del potencial electrostático generado con APBS, se muestran dos vistas: la cara que contiene los sitios activos con carácter básico y la cara opuesta con una zona central marcadamente ácida. Las cavidades catalíticas están resaltadas por dos circunferencias negras y las moléculas de PEG presentes en el cristal aparecen dibujadas como modelo de varillas.

#### III.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS

Con el fin de estudiar el mecanismo de reacción de la enzima y su función en la pared celular de la bacteria se llevaron a cabo una serie de experimentos de cristalización en los que se incubó la proteína con diversos sustratos y análogos estructurales del PG (Tabla III. 4). En el caso de BsrKo se realizaron

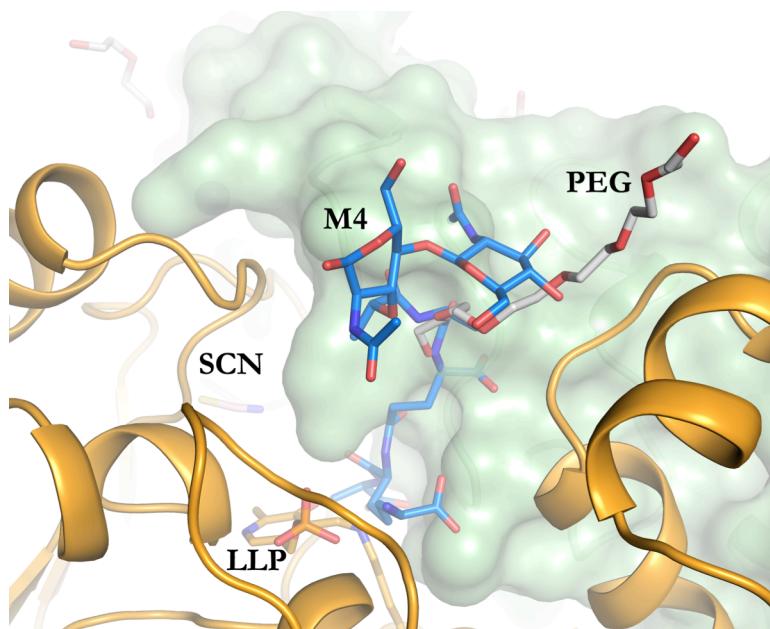
también los procedimientos comunes de cristalización en presencia de ligandos, la co-cristalización y la inmersión.

**Tabla III.4. Ligandos empleados para la cristalización con BsrKo**

Ligando	Fórmula química	Estructura	Tipo
<b>L-Met</b>	$C_5H_{11}NO_2S$		Aminoácido
<b>DHL-51</b>	$C_{32}H_{51}N_7O_{16}$		Análogo sintético de PG G(-)
<b>DHL-89</b>	$C_{40}H_{64}N_8O_{21}$		Análogo sintético de PG G(-)
<b>DHC-33</b>	$C_{15}H_{22}N_4O_6$		Análogo sintético de PG G(+)
<b>DHC-34</b>	$C_{17}H_{27}N_5O_7$		Análogo sintético de PG G(+)
<b>M4</b>	$C_{37}H_{63}N_7O_{21}$		Producido a partir de la degradación del PG de <i>E.coli</i>
<b>NAG</b>	$C_8H_{15}NO$		Unidad esencial de PG

La estructura de BsrKo presenta varias moléculas de PEG que están favoreciendo el empaquetamiento de la proteína en el cristal. Algunas de ellas se encontraban en la entrada a la cavidad del sitio activo (Fig. III.4b). Estas moléculas de PEG podrían estar simulando el complejo que formaría la proteína con las cadenas peptídicas que componen el PG. Hasta el momento se ha resuelto la estructura tridimensional de la proteína BsrKo en complejo con M4 (NAG-NAM-L-Ala-L-Glu-DAP-D-Ala), el

péptido entraría en la cavidad catalítica e interaccionaría directamente con el PLP mientras que el disacárido quedaría en la zona más externa de la cavidad. Al superponer la estructura de BsrKo en complejo con M4 con la BsrKo en complejo con PEG se observa que ambas moléculas se encuentran ocupando el mismo sitio en el canal de entrada al sitio activo (Fig. III.5).



**Figura III.5. Complejo cristalográfico de la racemasa BsrKo con el ligando M4, detalle de la entrada al sitio activo.** En colores se presentan cada uno de los monómeros que forman el sitio catalítico, en naranja el dominio N-terminal de un monómero y en superficie verde el dominio C-terminal del otro monómero. Los ligandos PEG y M4 se presentan en modelo de varillas blancas y azules, respectivamente. El LLP y el SCN (isotiocianato) aparecen coloreados en amarillo.

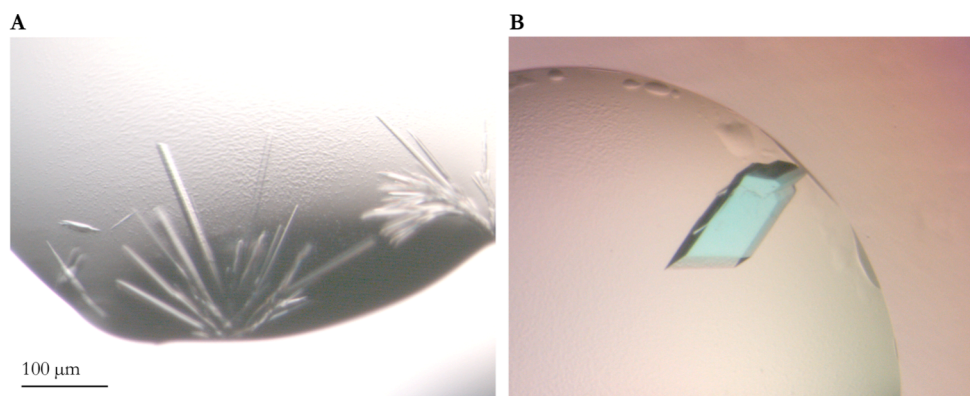
Además de los datos de difracción de BsrKo en complejo con M4 que se han mostrado, se han recogido otros datos con diferentes análogos de PG. Se está realizando un análisis exhaustivo de los mismos que permitirá entender mejor el mecanismo de acción de esta enzima. Los resultados obtenidos quedan recogidos en un artículo de investigación pendiente de ser publicados en una revista de ámbito científico.



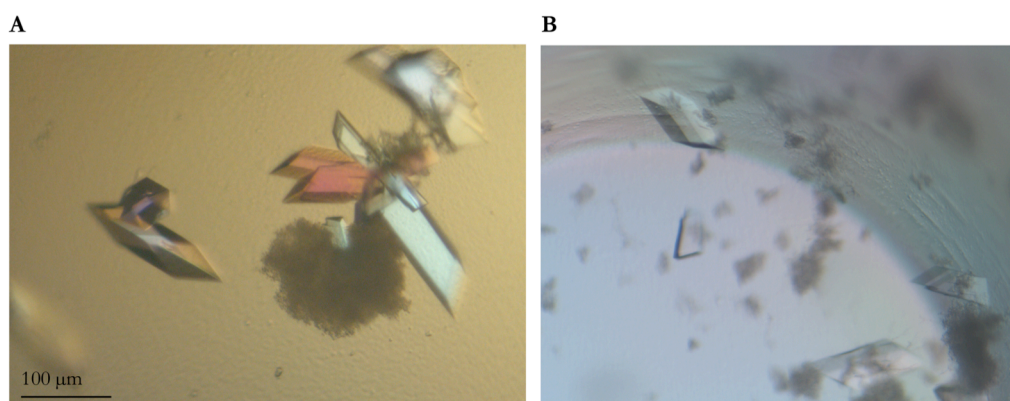
## 5.4. ANEXO IV. CRISTALIZACIÓN Y RESOLUCIÓN ESTRUCTURAL DE LA L<sub>1</sub>D-TRANSPEPTIDASA DE *E. coli* YnhG

### IV.I. CRISTALIZACIÓN

Tras el muestreo masivo de diferentes condiciones se obtuvieron varias condiciones de cristalización (Tabla IV.1). Se consiguieron cristales de gran tamaño que presentaban macla (Fig. IV.1).



**Figura IV.1. Cristales de YnhG.** a) Gota de cristalización inicial con cristales maclados. Condición de cristalización: 20% (m/v) PEG 8000, 0,2 M cloruro de magnesio y 0,1 M Tris pH 8,5. b) Cristal único de gran tamaño con macla obtenido en la condición de cristalización inicial 25% (m/v) PEG 3350, 0,2 M Acetato de amonio y 0,1 M HEPES pH 7,5.



**Figura IV.2. Cristales optimizados de YnhG.** a) Fotografía de cristales obtenidos en el muestreo de detergentes con IPTG. b) Cristales optimizados mediante la técnica de *microbatch*.

**Tabla IV.1. Condiciones de cristalización iniciales de YnhG**

Proteína	Precipitante	Tampón*	Aditivo*
<b>YnhG</b> 8 mg/ml en 20 mM Tris-HCl pH 7,5; 50 mM NaCl	20% (m/v) PEG 8000	Tris pH 8,5	MgCl <sub>2</sub>
	25% (m/v) PEG 3350	HEPES pH 7,5	Acetato amónico

\*Todos los tampones y aditivos se encuentran a 100 mM y 200 mM, respectivamente.



La optimización de los cristales obtenidos se llevó a cabo variando los porcentajes de PEG 3350 y el tampón HEPES empleando la técnica de difusión de vapor. Además, se realizó un muestreo de detergentes (Fig. IV.2 a). Los mejores cristales se forman a la semana en 25% (m/v) PEG 3350, 200 mM Acetato de amonio y 100 mM HEPES pH 7,5 bajo aceite (Fig. IV.2 b). Los cristales se congelaron directamente en nitrógeno líquido sin necesidad de añadir un agente crioprotector.

## IV.II. DETERMINACIÓN ESTRUCTURAL

La recogida de los datos de difracción se llevó a cabo en el sincrotrón ESRF (Grenoble) obteniendo un conjunto de datos a 2,2 Å de resolución. Los cristales pertenecen al sistema cristalino monoclinico, C2<sub>1</sub> con una dimensiones de celdilla de  $a = 151,6$  Å,  $b = 67,8$  Å,  $c = 155,8$  Å,  $\beta = 94,3$  °. En la Tabla IV.2 se recogen los resultados obtenidos del procesado de imágenes.

**Tabla IV.2. Estadísticas del procesado de datos de difracción.**

Parámetros	YnhG
Sistema cristalino	Monoclinico
Grupo espacial	C2 <sub>1</sub>
$a$ (Å)	151,6
$b$ (Å)	67,8
$c$ (Å)	155,8
$\alpha$ (°)	90
$\beta$ (°)	94,3
$\gamma$ (°)	90
Fuente de radiación	Sincrotrón
Longitud de onda (Å)	0,97626
Intervalo de resolución (Å)	77,70 – 2,22 (2,26 – 2,22)
Observaciones	441543
Reflexiones únicas	77763
Compleitud	98,9 (98,0)
Multiplicidad	5,7
* $R_{merge}$	0,09 (0,95)
$\langle I/\sigma(I) \rangle$	8,8 (1,7)

Los valores entre paréntesis corresponden a la resolución más alta.

\* $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - [I(hkl)]| / \sum_{hkl} \sum_i I_i(hkl)$ , donde  $I_i(hkl)$ .

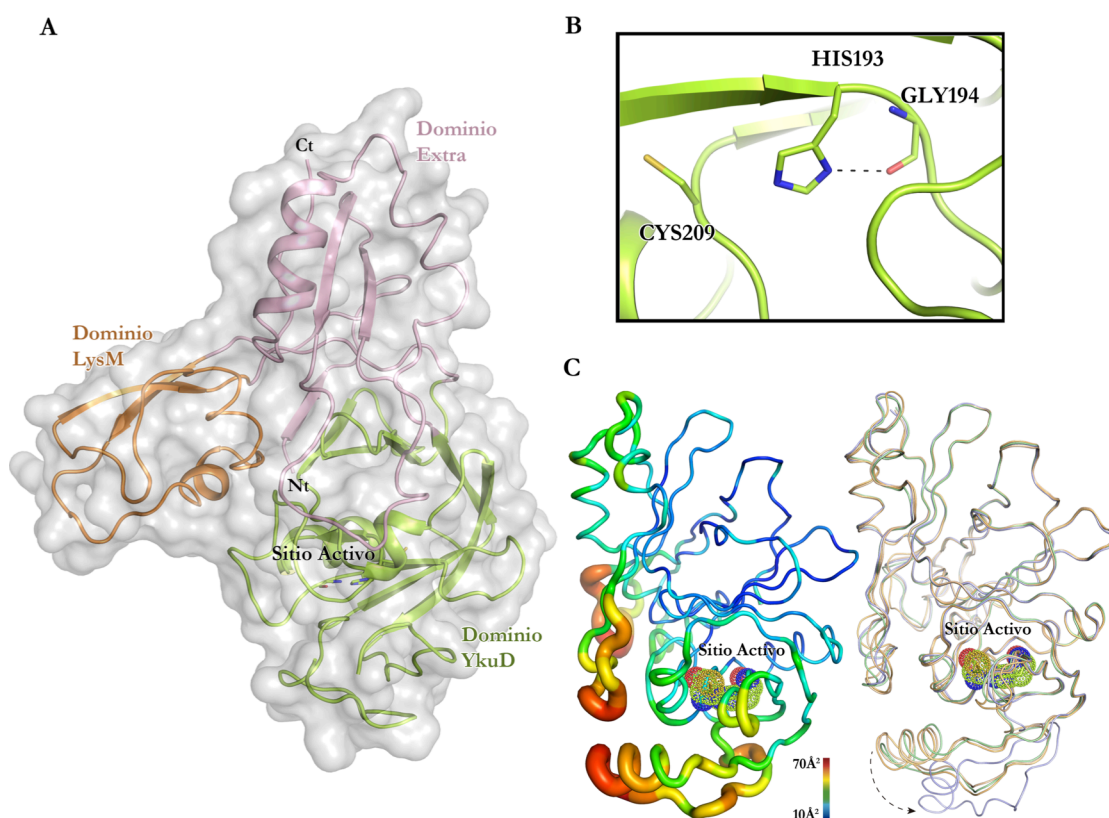
La estructura tridimensional se resolvió mediante el método de Remplazo Molecular usando el programa MOLREP con la L,D-transpeptidasa de *Klebsiella pneumoniae* (código PDB 4LZH) como modelo estructural, con la que presentaba un 49% de identidad, respectivamente. Mediante el cálculo del





## IV. III. ANÁLISIS ESTRUCTURAL

La proteína YnhG es una de las dos L,D-transpeptidasa que presenta *E. coli* que realizan los enlaces 3–3 entre los *meso*-DAP de dos cadena peptídicas adyacentes en el peptidoglicano. Como se describe en la introducción, los enlaces cruzados comunes en el PG son los 4–3 que establecen las enzimas D,D-transpeptidasas PBPs; sin embargo, los enlaces 3–3 aparecen sobre todo durante la fase estacionaria de crecimiento y confieren resistencia a la ampicilina. YnhG presenta tres dominios, un dominio tipo LysM de unión a PG, un dominio YkuD con actividad L,D-transpeptidasa donde se encuentra contenido el sitio activo y un dominio extra de función desconocida (Fig. IV.4 a). El sitio activo de la YnhG está formado por la His193; una cisteína, Cys209, se encarga de mantener la His en la conformación tautomérica correcta, además la cadena lateral de la Gly194 forma un enlace de hidrógeno con la His193 (Fig. IV.4 b).



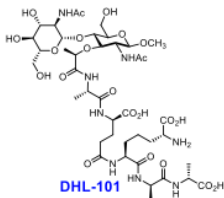
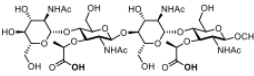
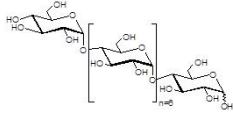
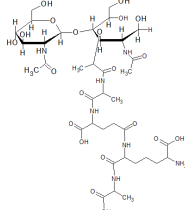
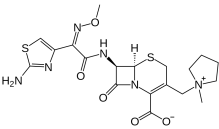
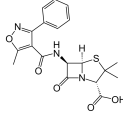
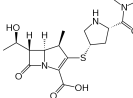
**Figura IV.4. Estructura general de YnhG y detalle del sitio activo.** a) YnhG presenta tres dominios diferenciados: un dominio LysM en naranja, un dominio transpeptidasa en verde y un dominio extra en rosa. b) Sitio activo de la L,D-transpeptidasa. Los residuos que intervienen en la reacción se muestran como varillas y el enlace de hidrógeno que se establece entre la Gly194 y la His193 aparece resaltado con una línea discontinua. c) A la izquierda, representación de los factores de agitación térmica para YnhG. Los valores cercanos al azul reflejan baja agitación térmica, mientras que si se acercan al rojo la agitación térmica es mayor lo que implica más flexibilidad. A la derecha se muestra una superposición de los cinco monómeros que componen el cristal de YnhG, con el bucle de uno de ellos en conformación más abierta. El sitio activo está representado mediante puntos.

Los diferentes dominios de la proteína presentan factores B medios bajos, por lo que la estructura en general es estable, no obstante contiene dos lazos cerca del sitio activo con factores de ocupación muy altos, indicando una mayor flexibilidad respecto del resto de la estructura; además, si se hace una superposición de los monómeros contenidos en el cristal de YnhG se puede apreciar que uno de ellos presenta el bucle en una conformación más abierta (Fig. IV.4 c).

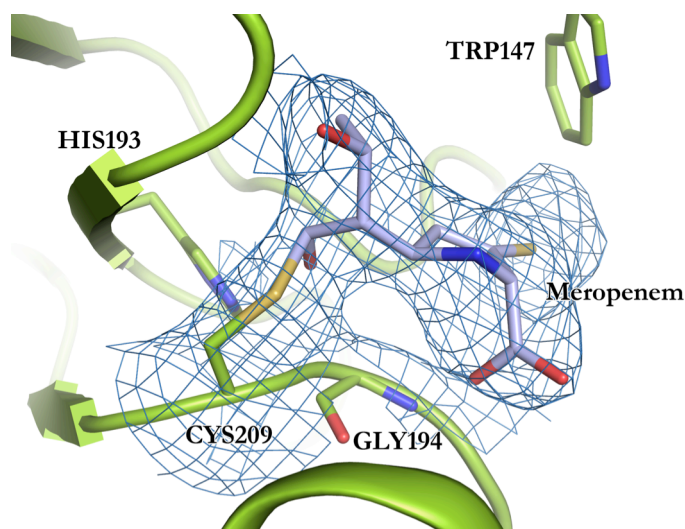
#### IV.IV. CRISTALIZACIÓN EN PRESENCIA DE SUSTRATOS

Se han llevado a cabo diversos experimentos de co-cristalización en presencia de sustratos que inhiben la actividad de la enzima, como son las carbapenemas (meropenem) y antibióticos  $\beta$ -lactámicos. Además, para entender la función del dominio LysM, se realizaron experimentos de co-cristalización en presencia de análogos de PG y de maltooctaosa. En la tabla IV.4 se listan los diferentes ligandos empleados para la cristalización en complejo con YnhG.

Tabla IV.4. Ligandos empleados para la cristalización con YnhG

Ligando	Fórmula química	Estructura	Tipo
DHL-101	$C_{41}H_{68}N_8O_{22}$		Análogo sintético de PG G(-)
DH-166	$C_{39}H_{64}N_4O_{25}$		Análogo sintético de PG G(-)
Maltooctaosa	$C_{48}H_{82}O_{41}$		Polisacárido de 8 azúcares de glucosa
M4	$C_{37}H_{63}N_7O_{21}$		Producido a partir de la degradación del PG de <i>E.coli</i>
Cefepima	$C_{19}H_{24}N_6O_5S_2$		Antibiótico del grupo de las cefalosporinas
Oxacilina	$C_{19}H_{19}N_3O_5S_2$		Antibiótico $\beta$ -lactámico
Meropenem	$C_{17}H_{25}N_3O_5S_2$		Antibiótico $\beta$ -lactámico

Se han obtenido datos de difracción de YnhG en complejo con Meropenem, el antibiótico forma un enlace covalente con la Cys209 (Fig. IV.5). Por otro lado, el mapa de densidad correspondiente al conjunto de datos de YnhG con Meropenem no permitió trazar la cadena polipeptídica correspondiente al bucle que presenta movilidad cerca del sitio activo confirmando la flexibilidad del bucle. Además, se han obtenido varios cristales de YnhG por co-cristalización con diferentes análogos de PG, así como con el polisacárido.



**Figura IV.5. Complejo cristalográfico de YnhG con el antibiótico Meropenem.** Se representa el dominio que contiene el sitio activo en verde, los residuos involucrados en la interacción están resaltados en modelo de varillas. El meropenem (en azul) se encuentra unido covalentemente a la Cys209. La densidad electrónica corresponde al mapa  $2F_o - F_c$  a  $0,8\sigma$ .

Actualmente se está analizando el conjunto de datos de difracción de YnhG en complejo con los diferentes ligandos. Estos resultados servirán para responder a la siguientes cuestiones: cómo se produce la inhibición de la actividad enzimática de la proteína, cómo realiza la actividad L,D-transpeptidasa y, también, cuál es la localización de la enzima en el PG de *E. coli*.

## 6. CONSIDERACIONES FINALES



En este trabajo se ha realizado la caracterización estructural y funcional de dos familias de proteínas de superficie implicadas en la regulación de la metabolismo de la pared celular bacteria y en interacciones con proteínas de la matriz extracelular del hospedador. A continuación se resumen los resultados más importantes para cada una de ellas y sus implicaciones en la fisiología de la bacteria.

### **6.1. EL PAPEL MULTIFUNCIONAL DE LA FOSFOGLICERATO QUINASA DE NEUMOCOCO Y SU IMPLICACIÓN EN VIRULENCIA**

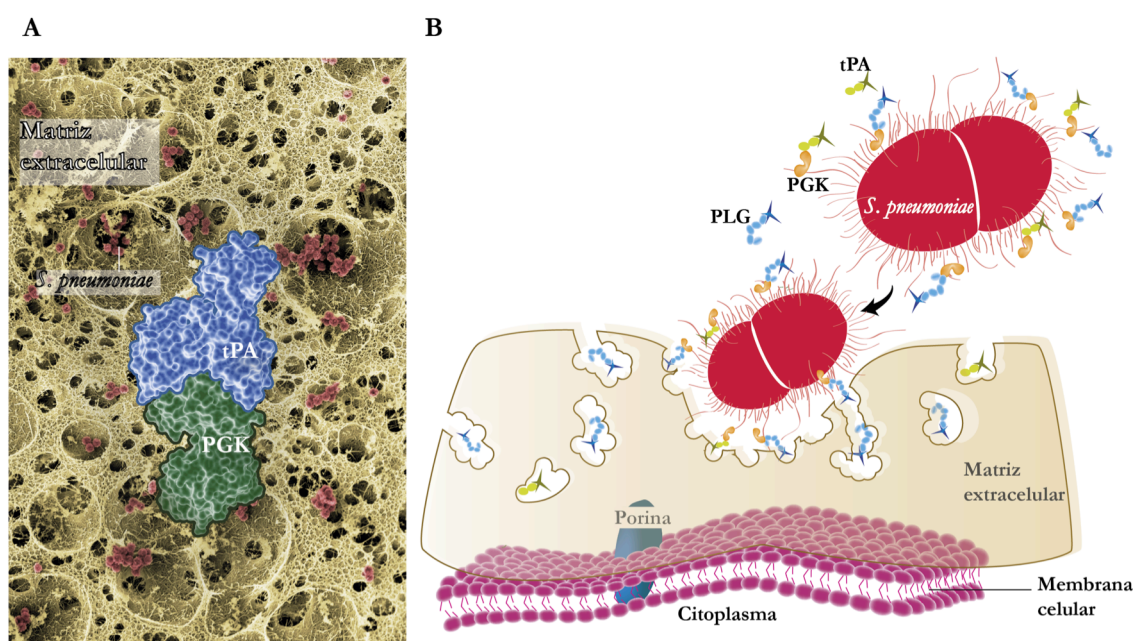
El papel de la proteína de neumococo PGK en la glicólisis es ampliamente conocido; sin embargo, recientemente hemos demostrado que, además de participar en el metabolismo celular, se encuentra localizada en la superficie celular. Este hecho ha permitido clasificarla dentro del grupo de proteínas no clásicas de superficie (NCSP), ya que son proteínas que carecen de péptido líder o motivos de anclaje a membrana.

*Streptococcus pneumoniae* es capaz de unir plasminógeno (PLG) humano a través de las proteínas que se encuentran expuesta en la superficie del patógeno. El reclutamiento de PLG en la superficie del neumococo promueve la conversión del PLG en plasmina (serín proteasa), favoreciendo la disolución de las cadenas de trombina y la degradación de la matriz extracelular del hospedador. Este mecanismo de interacción aumenta la virulencia del patógeno ya que facilita su diseminación a través de los tejidos del hospedador. A través de ensayos de microscopía electrónica se ha podido determinar la localización de la PGK en la superficie del neumococo y junto con el PLG, confirmando la interacción de la proteína PGK con el PLG.

La determinación estructural de la PGK de neumococo ha permitido determinar que el sitio de unión a PLG se encuentra localizado en el dominio N-terminal de la proteína en la cara opuesta al sitio activo, de manera que la actividad enzimática no se vería afectada.

Se sabe que el PLG interacciona con otras proteínas a través de los sitios de unión a lisina que contienen sus dominios *kringle*. El sitio de unión a lisina está formado por las cadenas laterales de dos triptófanos y varios residuos cargados alrededor. La proteína PGK de neumococo presenta una zona de lisinas expuestas que podrían mediar la interacción con el PLG. El modelo de interacción de PGK con el dominio K2 del PLG generado a partir de la estructura cristalográfica de PGK ha permitido establecer las bases de la interacción. Recientemente se ha resuelto la estructura del PLG humano el cual presenta las mismas interacciones en el K2 y el dominio serín proteasa. El modelo de interacción entre la PGK y el K2 propuesto planteó la posibilidad de que la PGK pudiera interaccionar con otras proteínas que presentaran dominios *kringle*, como son los activadores de PLG (uPA y tPA). Además, mediante ensayos bioquímicos se ha confirmado que uno de los activadores (tPA) interacciona con la PGK de neumococo. Hasta la fecha, es la primera vez que se describe la interacción de un activador de PLG con *S. pneumoniae*.

La interacción con los dominios *kringle* del extremo N-terminal produciría una apertura conformacional que activaría el centro proteolítico del PLG. Los resultados del análisis de la actividad de la plasmina indican que la unión del PLG a la PGK promovería la conversión del PLG en plasmina a través de los activadores tPA o uPA. Una propiedad única que presenta la PGK de neumococo es la capacidad de reclutar el tPA, activando el plasminógeno y la degradación de las cadenas de fibrina de manera eficaz. La capacidad de unir tanto PLG como tPA en la superficie del neumococo favorecería la conversión del PLG en plasmina a través del tPA en la zona donde se está produciendo la infección, potenciando la degradación de la matriz extracelular y facilitando la dispersión del patógeno dentro del hospedador (Fig. 4.1).



**Figura 6.1. Interacción de PGK con proteínas del hospedador y su implicación en patogénesis. a)** Fotografía de microscopía electrónica en la que se ve cómo el neumococo degrada activamente la matriz extracelular, el neumococo aparece coloreado en rojo y la matriz extracelular en amarillo. Sobre la fotografía se muestra el modelo estructural del complejo entre la proteína de neumococo PGK (superficie verde) y la angiostatina (ANG) (superficie azul) generado a partir de la estructura cristalográfica de la PGK. **b)** Representación esquemática del reclutamiento de plasminógeno y tPA en la superficie del neumococo. La PGK está expuesta en la superficie del neumococo pudiendo reclutar tanto el plasminógeno como la tPA. El efecto combinado de ambas proteínas es empleado por la bacteria para degradar la matriz extracelular del hospedador permitiendo así la diseminación y dispersión del patógeno a través de los tejidos.

Dado que la fibrinólisis está relacionada con los procesos inflamatorios, nuestros resultados describirían un nuevo escenario en el que el neumococo desestabilizaría la maquinaria de fibrinólisis del hospedador durante la fase inicial de colonización. Este hecho potenciaría tanto la diseminación a través de los tejidos como la entrada en el sistema circulatorio.



## 6.2. MAQUINARIA CATALÍTICA E IMPLICACIONES *IN VIVO* DE UNA NUEVA FAMILIA DE RACEMASAS DE AMPLIO ESPECTRO.

Se ha identificado una nueva familia de racemasas, las racemasas de amplio espectro (BsR), capaces de generar D-aminoácidos no canónicos (NCDAA). Estas Bsrs se encuentran localizadas en el espacio periplásmico de la bacteria, en contraposición a las racemasas mono específicas (AlaR) presentes en el citoplasma celular. Los NCDAA están involucrados en gran variedad de procesos celulares, tales como la modulación de la estructura de la pared celular, la esporulación y la disolución de biofilms, entre otros.

Nuestro análisis exhaustivo de un miembro representativo de esta familia de racemasas de amplio espectro, la racemasa de *Vibrio cholerae* (BsRv), a través de diversas técnicas de cristalografía de rayos X, bioquímicas y bioinformáticas ha permitido identificar los atributos que determinan la multiespecificidad de dicha enzima.

Hemos identificado una serie de residuos conservados que definen la huella molecular del sitio activo, permitiendo diferenciar las racemasas de amplio espectro de las alanina racemasas (AlaR). Además, los residuos que conforman la huella molecular están directamente relacionados con la multiespecificidad de BsRv. El análisis estructural de otra BsR permitió a su vez confirmar que los factores que determinan la plasticidad de BsRv están conservados en los miembros de la familia de racemasas de amplio espectro.

Aunque inicialmente se pensaba que BsRv únicamente era capaz de producir los aminoácidos no canónicos D-Met y D-Leu, en este trabajo hemos demostrado que esta enzima puede racemizar ocho aminoácidos proteinogénicos más, así como algunos aminoácidos no proteinogénicos. Los sustratos de la enzima se pueden dividir en dos clases: los aminoácidos alifáticos que no están ramificados en  $\beta$  (p. ej. Ala, Ser, Leu, Cys, Gln, Asn y Met) y los aminoácidos básicos (p. ej. Arg, Lys e His).

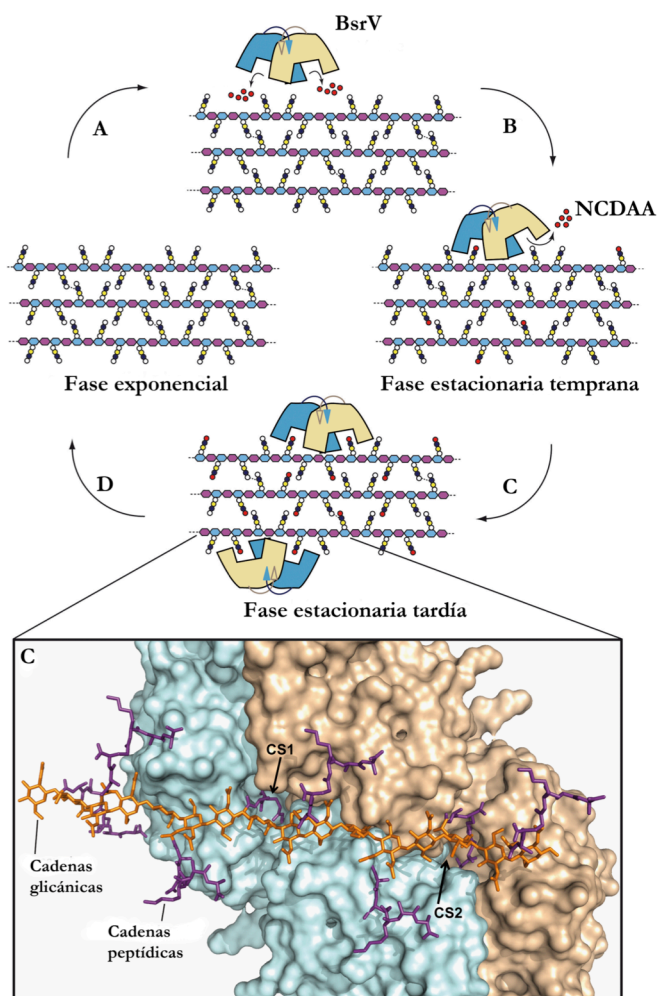
Un análisis comparativo de las estructuras de miembros de la familia de las BsR y de las AlaR permitió determinar las características esenciales que determinan la plasticidad de estas racemasas. Estos determinantes estructurales están relacionados con la mayor amplitud de la cavidad catalítica así como el entorno catalítico que permite la entrada de sustratos de muy diversa naturaleza química. Además, análisis estructurales, bioquímicos y de modelado sugieren que el canal de entrada de la BsRv podría acomodar cadenas peptídicas.

Se ha propuesto un modelo de acción de la BsRv *in vivo* que explicaría la regulación del peptidoglicano mediada por NCDAA en *Vibrio cholerae*. Durante la fase exponencial de crecimiento de la bacteria BsRv no se expresa y el PG está compuesto únicamente de aminoácidos canónicos (D-Ala y D-Glu). Una vez que la bacteria entra en la fase estacionaria, la BsRv se exporta al espacio periplásmico donde se acumulan altas concentraciones de NCDAA. El PG incorpora estos NCDAA inhibiendo así la actividad de la BsRv y evitando una modificación abusiva del peptidoglicano y una excesiva acumulación



de aminoácidos no canónicos libres. BsrV deja de producirse cuando el crecimiento de la bacteria finaliza (Fig. 4.2). La producción controlada de NCDAA por parte de la bacteria podría ser crucial para la adaptación de la pared celular frente a condiciones perjudiciales.

Teniendo en cuenta la estructura cristalográfica de la proteína BsrV proponemos un modelo de interacción de la proteína con el modelo de panal de abeja del peptidoglicano (Meroueh *et al.*, 2006) donde los dos sitios activos de BsrV encajarían en las cadenas peptídicas del PG, demostrando así el



**Figura 6.2.** Esquema de la regulación que ejerce la BsrV sobre el PG mediada por NCDAA. **a)** En la fase exponencial de crecimiento, *V. Cholerae* no expresa BsrV y el PG presenta únicamente aminoácidos canónicos. **b)** En la fase estacionaria la BsrV produce altas concentraciones de NCDAA que se acumulan en el periplasma y el PG contiene modificaciones. **c)** Las modificaciones presentes en el PG provocan la inhibición de la proteína BsrV, evitando una acumulación excesiva de NCDAA libres. En la ampliación se muestra el modelo de interacción generado a partir de la estructura tridimensional de BsrV (en superficie amarilla y azul) y el modelo de panal de abeja del PG (representado como modelo de varillas, en naranja las cadenas glicánicas y en morado las peptídicas), los sitios activos de la enzima aparecen como CS1 y CS2. **d)** La racemasa BsrV deja de expresarse cuando el crecimiento bacteriano finaliza, no hay producción de NCDAA ni la consiguiente incorporación en el PG.

bloqueo de la enzima en la etapa final de crecimiento bacteriano (Fig. 4.2).

Dado el impacto que tienen los NCDAA en una gran variedad de procesos celulares, tales como la estabilidad del biofilm, la esporulación celular, la comunicación celular y la síntesis del peptidoglicano, entre otras; tanto el análisis estructural como el mecanismo de regulación de esta familia de racemasas de amplio espectro abre una nueva vía para el diseño de drogas dirigidas a dificultar la adaptabilidad de la bacteria.

## 7. CONCLUSIONES



A continuación se enumeran las principales conclusiones extraídas de este trabajo de investigación:

1. Se ha cristalizado y resuelto mediante cristalografía de rayos X la estructura tridimensional de la proteína PGK de *S. pneumoniae*. La proteína se compone de dos dominios globulares, el dominio N-terminal contiene el sitio de unión al sustrato y el dominio C-terminal el sitio catalítico.
2. Experimentos de microscopía electrónica demostraron que la proteína PGK está localizada en la superficie celular de neumococo. La PGK carece de un motivo de anclaje a membrana y se desconoce el mecanismo de exportación al exterior celular, por lo que se la ha clasificado dentro de la familia de proteínas no clásica de superficie (NCSP).
3. Se ha demostrado que la PGK de neumococo es capaz de unir tanto plasminógeno (PLG) como angiostatina (ANG). El análisis estructural de la proteína PGK ha determinado que el sitio de unión a plasminógeno (PLG) se encuentra en el dominio N-terminal de la proteína, en la cara opuesta al sitio activo.
4. El modelo computacional basado en la estructura cristalográfica de PGK en complejo con el *kringle 2* (K2) del plasminógeno (PLG) permitió describir los detalles de la interacción. Este modelo sugirió la potencial interacción de la PGK con los activadores de PLG que contenían dominios *kringle*.
5. La PGK es capaz de unir el activador de plasminógeno tPA pero no el uPA, el análisis estructural reveló cambios importantes en el sitio de unión a lisina que presentaba uPA que impedían la interacción con PGK.
6. Se han definido los diferentes tipos de enzimas PLP-dependientes atendiendo a su plegamiento estructural.
7. Se ha identificado una nueva familia de racemasas de amplio espectro dentro de la cual se encuentra la racemasa de *Vibrio cholerae* BsrV.
8. Se ha descrito el mecanismo catalítico de las alanina racemasas (AlaR) y comparado con el de las racemasas de amplio espectro (BsrV). La caracterización estructural y los estudios bioinformáticos han permitido explicar las diferencias en la selectividad de sustrato entre BsR y AlaR.

9. Se ha resuelto mediante cristalografía de rayos X la racemasa de amplio BsrV de *V. cholerae* a resolución atómica. La enzima BsrV presenta un plegamiento globular, está constituida por dos dominios, el extremo amino contiene el sitio activo.
10. Se ha resuelto la estructura cristalográfica de la racemasa de amplio espectro Bsr<sub>Ah</sub> de *Aeromonas hydrophila*.
11. Se han establecido las bases estructurales que determinan una racemasa de amplio espectro. Además, se ha descrito la huella molecular que permitirá identificar nuevas enzimas de amplio espectro.
12. Se han concretado los determinantes esenciales que permiten la plasticidad del mecanismo catalítico de las racemasas de amplio espectro, relativos a la disposición del sitio de entrada, la amplitud del canal y el entorno catalítico.

A continuación se muestran las conclusiones relativas a los trabajos de investigación pendientes de ser publicados:

13. Se ha cristalizado y resuelto mediante cristalografía de rayos X la racemasa de *Bacillus subtilis* Bsu17640. La estructura de la proteína presenta un plegamiento globular con dos dominios diferenciados, el dominio amino terminal contiene el sitio activo. El sitio activo presenta dos moléculas de TRIS que se encuentran estabilizadas por interacciones de hidrógeno y polares.
14. Se ha cristalizado y resuelto la estructura de la racemasa de amplio espectro de *Acinetobacter baumannii* BsrAb. La proteína presenta una estructura homodimérica con un plegamiento globular con dos dominios diferenciados, amino terminal y carboxilo terminal. La proteína contiene dos lazos flexibles en la zona que conecta los dos sitios de unión. Se ha resuelto la estructura de BsrAb en complejo con el aminoácido L-Met.
15. Se ha cristalizado y determinado la estructura tridimensional de la racemasa de amplio espectro de *Kingella oralis* BsrKo. La proteína presenta el plegamiento típico de las racemasas PLP dependientes con dos dominios amino y carboxilo terminal diferenciados. Se han obtenido cristales de BsrKo en complejo con la unidad estructural de peptidoglicano NAG-NAM-L-Ala-D-Glu-L,D-Aminopimérico-D-Ala (M4).
16. Se ha cristalizado y resuelto la estructura tridimensional de la L,D-transpeptidasa de *E. coli* YnhG. La proteína presenta tres dominios, un dominio LysM, un dominio YkuD y un dominio extra. El plegamiento de YnhG es globular, se aprecian dos lazos flexibles rodeando al sitio activo. Se ha resuelto la estructura de YnhG en complejo con el antibiótico Meropenem y se han obtenido varios cristales por co-cristalización con análogos de peptidoglicano.

## 8. BIBLIOGRAFÍA



- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.
- Aguilar, C., Vlamakis, H., Losick, R. & Kolter, R. (2007). *Curr Opin Microbiol* **10**, 638-643.
- Altschul, S. F. & Lipman, D. J. (1990). *Proceedings of the National Academy of Sciences of the United States of America* **87**, 5509-5513.
- Austrian, R. (1981). *Proceedings of the American Philosophical Society* **125**, 46-51.
- Baley, S. (1994). *Acta Cryst. D* **50**, 157-163.
- Bergmann, S. & Hammerschmidt, S. (2006). *Microbiology* **152**, 295-303.
- Bergmann, S. & Hammerschmidt, S. (2007). *Thromb Haemost* **98**, 512-520.
- Bergmann, S., Rohde, M., Chhatwal, G. S. & Hammerschmidt, S. (2001). *Molecular microbiology* **40**, 1273-1287.
- Bergmann, S., Rohde, M., Preissner, K. T. & Hammerschmidt, S. (2005). *Thromb Haemost* **94**, 304-311.
- Blasco, B., Pisabarro, A. G. & de Pedro, M. A. (1988). *J Bacteriol* **170**, 5224-5228.
- Bodanszky, M. & Perlman, D. (1969). *Science* **163**, 352-358.
- Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. (2006). *Mol Microbiol* **59**, 1229-1238.
- Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. (2001). *Proc Natl Acad Sci U S A* **98**, 11621-11626.
- Carneiro, C. R., Postol, E., Nomizo, R., Reis, L. F. & Brentani, R. R. (2004). *Microbes Infect* **6**, 604-608.
- Cartwright, K. (2002). *European journal of pediatrics* **161**, 188-195.
- Castellino, F. J. & Powell, J. R. (1981). *Methods Enzymol* **80 Pt C**, 365-378.
- Cava, F., Lam, H., de Pedro, M. A. & Waldor, M. K. (2011). *Cell Mol Life Sci* **68**, 817-831.
- Chavakis, T., Athanasopoulos, A., Rhee, J. S., Orlova, V., Schmidt-Woll, T., Bierhaus, A., May, A. E., Celik, I., Nawroth, P. P. & Preissner, K. T. (2005). *Blood* **105**, 1036-1043.
- Chayen, N. E. (1998). *Acta crystallographica. Section D, Biological crystallography* **54**, 8-15.
- Cortes-Cabrera, A., Gago, F. & Morreale, A. (2015). *Methods Mol Biol* **1289**, 89-100.



- Counago, R. M., Davlieva, M., Strych, U., Hill, R. E. & Krause, K. L. (2009). *BMC structural biology* **9**, 53.
- DeLano, W. L. (2002). *Current opinion in structural biology* **12**, 14-20.
- Donofrio, F. C., Calil, A. C., Miranda, E. T., Almeida, A. M., Benard, G., Soares, C. P., Veloso, S. N., Soares, C. M. & Mendes Giannini, M. J. (2009). *J Med Microbiol* **58**, 706-713.
- Driehuis, F. & Wouters, J. T. (1987). *J Bacteriol* **169**, 97-101.
- Driks, A. (1999). *Microbiol Mol Biol Rev* **63**, 1-20.
- Duboscq, C., Genoud, V., Parborell, M. F. & Kordich, L. C. (1997). *Thromb Res* **86**, 505-513.
- Dunlop, D. S., Neidle, A., McHale, D., Dunlop, D. M. & Lajtha, A. (1986). *Biochem Biophys Res Commun* **141**, 27-32.
- Eliot, A. C. & Kirsch, J. F. (2004). *Annu Rev Biochem* **73**, 383-415.
- Emsley, P. & Cowtan, K. (2004). *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132.
- Espaillet, A., Carrasco-Lopez, C., Bernardo-Garcia, N., Pietrosevoli, N., Otero, L. H., Alvarez, L., de Pedro, M. A., Pazos, F., Davis, B. M., Waldor, M. K., Hermoso, J. A. & Cava, F. (2014). *Acta Crystallogr D Biol Crystallogr* **70**, 79-90.
- Fisher, G. H., D'Aniello, A., Vetere, A., Padula, L., Cusano, G. P. & Man, E. H. (1991). *Brain Res Bull* **26**, 983-985.
- Flack, H. D. (2009). *Acta Crystallogr A* **65**, 371-389.
- Fuchs, S. A., Berger, R., Klomp, L. W. & de Koning, T. J. (2005). *Mol Genet Metab* **85**, 168-180.
- Gal, J. (2008). *Chirality* **20**, 5-19.
- Garcia, J. L., Sanchez-Beato, A. R., Medrano, F. J. & Lopez, R. (1998). *Microbial drug resistance* **4**, 25-36.
- Grishin, N. V., Phillips, M. A. & Goldsmith, E. J. (1995). *Protein Sci* **4**, 1291-1304.
- Halvorson, H. O. & Spiegelman, S. (1952). *J Bacteriol* **64**, 207-221.
- Hashimoto, A., Nishikawa, T., Oka, T. & Takahashi, K. (1993). *J Neurochem* **60**, 783-786.
- Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., 3rd, Deng, S. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R. & Shewchuk, L. M. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 72-79.
- Hellberg, J. E., Matilla, M. A. & Salmond, G. P. (2015). *Front Microbiol* **6**, 137.

- Henrichsen, J. (1995). *Journal of clinical microbiology* **33**, 2759-2762.
- Hills, G. M. (1949). *Biochem J* **45**, 353-362.
- Holm, L. & Rosenstrom, P. (2010). *Nucleic Acids Res* **38**, W545-549.
- Holtje, J. V. (1998). *Microbiol Mol Biol Rev* **62**, 181-203.
- Homoe, P., Bjarnsholt, T., Wessman, M., Sorensen, H. C. & Johansen, H. K. (2009). *Eur Arch Otorhinolaryngol* **266**, 1533-1538.
- Hubbard, M. A., Thorkildson, P., Welch, W. H. & Kozel, T. R. (2013). *Mol Immunol* **55**, 337-344.
- Hughes, M. J., Moore, J. C., Lane, J. D., Wilson, R., Pribul, P. K., Younes, Z. N., Dobson, R. J., Everest, P., Reason, A. J., Redfern, J. M., Greer, F. M., Paxton, T., Panico, M., Morris, H. R., Feldman, R. G. & Santangelo, J. D. (2002). *Infection and immunity* **70**, 1254-1259.
- Jacobsen, S. M., Stickler, D. J., Mobley, H. L. & Shirliff, M. E. (2008). *Clin Microbiol Rev* **21**, 26-59.
- Jayaraman, A. & Wood, T. K. (2008). *Annu Rev Biomed Eng* **10**, 145-167.
- Kaku, H., Nishizawa, Y., Ishii-Minami, N., Akimoto-Tomiyama, C., Dohmae, N., Takio, K., Minami, E. & Shibuya, N. (2006). *Proc Natl Acad Sci U S A* **103**, 11086-11091.
- Karatan, E. & Watnick, P. (2009). *Microbiol Mol Biol Rev* **73**, 310-347.
- Kassar, R., Hachem, R., Jiang, Y., Chaftari, A. M. & Raad, I. (2009). *Medicine (Baltimore)* **88**, 279-283.
- Kleckner, N. W. & Dingle, R. (1988). *Science* **241**, 835-837.
- Kobayashi, H. (2005). *Treat Respir Med* **4**, 241-253.
- Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R. & Losick, R. (2010). *Science* **328**, 627-629.
- Lam, H., Oh, D. C., Cava, F., Takacs, C. N., Clardy, J., de Pedro, M. A. & Waldor, M. K. (2009). *Science* **325**, 1552-1555.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). *J Biomol NMR* **8**, 477-486.
- Liu, Y. C. & Post, J. C. (2009). *Curr Allergy Asthma Rep* **9**, 449-455.
- Magnet, S., Bellais, S., Dubost, L., Fourgeaud, M., Mainardi, J. L., Petit-Frere, S., Marie, A., Mengin-Lecreulx, D., Arthur, M. & Gutmann, L. (2007). *J Bacteriol* **189**, 3927-3931.
- Matthews, B. W. (1968). *J Mol Biol* **33**, 491-497.

- McKevitt, M. T., Bryant, K. M., Shakir, S. M., Larabee, J. L., Blanke, S. R., Lovchik, J., Lyons, C. R. & Ballard, J. D. (2007). *Infect Immun* **75**, 5726-5734.
- Meierhenrich, U. (2008). *Amino Acids and the Asymmetry of Life*, pp. 17-46: Springer Berlin Heidelberg.
- Meroueh, S. O., Bencze, K. Z., Hesek, D., Lee, M., Fisher, J. F., Stemmler, T. L. & Mobashery, S. (2006). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4404-4409.
- Miller, M. B. & Bassler, B. L. (2001). *Annu Rev Microbiol* **55**, 165-199.
- Miyashita, C., Wenzel, E. & Heiden, M. (1988). *Haemostasis* **18 Suppl 1**, 7-13.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta crystallographica. Section D, Biological crystallography* **53**, 240-255.
- Pastoret, S., Fraipont, C., den Blaauwen, T., Wolf, B., Aarsman, M. E., Piette, A., Thomas, A., Brasseur, R. & Nguyen-Disteche, M. (2004). *Journal of bacteriology* **186**, 8370-8379.
- Peltier, J., Courtin, P., El Meouche, I., Lemee, L., Chapot-Chartier, M. P. & Pons, J. L. (2011). *J Biol Chem* **286**, 29053-29062.
- Perez-Dorado, I., Galan-Bartual, S. & Hermoso, J. A. (2012). *Mol Oral Microbiol* **27**, 221-245.
- Piatigorsky, J. & Wistow, G. J. (1989). *Cell* **57**, 197-199.
- Pisabarro, A. G., de Pedro, M. A. & Vazquez, D. (1985). *J Bacteriol* **161**, 238-242.
- Reslinski, A., Mikucka, A., Szmytkowski, J., Gospodarek, E. & Dabrowiecki, S. (2009). *Pol J Microbiol* **58**, 367-369.
- Rodriguez-Bano, J., Marti, S., Soto, S., Fernandez-Cuenca, F., Cisneros, J. M., Pachon, J., Pascual, A., Martinez-Martinez, L., McQueary, C., Actis, L. A., Vila, J. & Spanish Group for the Study of Nosocomial, I. (2008). *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **14**, 276-278.
- Romero, D., Aguilar, C., Losick, R. & Kolter, R. (2010). *Proc Natl Acad Sci U S A* **107**, 2230-2234.
- Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. (2008). *FEMS Microbiol Rev* **32**, 234-258.
- Sengupta, S., Ghosh, S. & Nagaraja, V. (2008). *Microbiology* **154**, 2796-2803.
- Setlow, P. (2003). *Curr Opin Microbiol* **6**, 550-556.
- Severi, E., Hood, D. W. & Thomas, G. H. (2007). *Microbiology* **153**, 2817-2822.
- Soda, K., Yoshimura, T. & Esaki, N. (2001). *Chem Rec* **1**, 373-384.

- Stragier, P. & Losick, R. (1996). *Annual review of genetics* **30**, 297-241.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res* **22**, 4673-4680.
- Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. (2008). *Genes Dev* **22**, 945-953.
- Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. (2013). *Nature reviews. Microbiology* **11**, 157-168.
- Vollmer, W. (2007). *Structure and Biosynthesis of the Pneumococcal Cell Wall*, Norfolk: Horizon Bioscience ed.
- Winkler, F., Kastenbauer, S., Koedel, U. & Pfister, H. W. (2002). *Neurology* **59**, 1350-1355.
- Wolosker, H. (2007). *Mol Neurobiol* **36**, 152-164.
- Wolosker, H., Panizzutti, R. & De Miranda, J. (2002). *Neurochem Int* **41**, 327-332.
- Wu, H. M., Kuan, Y. C., Chu, C. H., Hsu, W. H. & Wang, W. C. (2012). *PloS one* **7**, e48301.
- Yang, H., Wang, M., Yu, J. & Wei, H. (2015). *FEMS Microbiol Lett* **362**.
- Yankah, A. C., Pasic, M., Klose, H., Siniawski, H., Weng, Y. & Hetzer, R. (2005). *Eur J Cardiothorac Surg* **28**, 69-75.
- Yother, J. (2004). *Capsules*.

